

LOCALIZATION IN RANDOM GEOMETRIC GRAPHS WITH TOO MANY EDGES

SOURAV CHATTERJEE AND MATAN HAREL

ABSTRACT. Consider a random geometric graph $G(\chi_n, r_n)$, given by connecting two vertices of a Poisson point process χ_n of intensity n on the unit torus whenever their distance is smaller than the parameter r_n . The model is conditioned on the rare event that the number of edges observed, $|E|$, is greater than $(1 + \delta)\mathbb{E}(|E|)$, for some fixed $\delta > 0$. This article proves that upon conditioning, with high probability there exists a ball of diameter r_n which contains a clique of at least $\sqrt{2\delta\mathbb{E}(|E|)}(1 - \varepsilon)$ vertices, for any $\varepsilon > 0$. Intuitively, this region contains all the “excess” edges the graph is forced to contain by the conditioning event, up to lower order corrections. As a consequence of this result, we prove a large deviations principle for the upper tail of the edge count of the random geometric graph.

1. INTRODUCTION

The Random Geometric Graph is a simple stochastic model, first studied in [13] in 1972, for generating a graph: given the parameters n and r , consider a Poisson point process of intensity n on the unit torus, equipped with some norm $\|\cdot\|$, and declare an edge between any two vertices that are distance $\leq r$ from each other.

Unlike the well-known Erdős–Rényi random graph, the random geometric graph’s definition leads to strong dependence between edges: if three vertices form a “V” shaped graph, they are far more likely to have the third edge of the triangle than if no assumption were made on the other edges, as a consequence of the triangle inequality.

Many properties of this graph model have been studied. The classic monograph of Mathew Penrose [23] studies results pertaining to many graph-theoretical functions of random geometric graphs, including (but not limited to) laws of large numbers and central limit theorems for subgraph counts, independence number, and chromatic number, as well as many properties connected to the giant component. Many of the results presented in this monograph have been improved and generalized by Penrose and his coauthors in the 12 years since its initial publication. Besides this, there have

2010 *Mathematics Subject Classification.* 60F10, 05C80, 60D05.

Key words and phrases. Random geometric graph, Poisson point process, large deviation, localization.

Research partially supported by ERC AG “COMPASP” and NSF grant DMS-1005312.

been investigations into other probabilistic features, such as threshold functions for cover times and mixing times [2] and a characterization of sharp thresholds for many monotone graph functions [11]. This list is far from comprehensive, of course, and the random geometric graph is still an active research area.

The random geometric graph is also highly related to the random connection continuum percolation model. In that model, the vertex set is given by an (almost surely infinite) Poisson point process of fixed intensity on \mathbb{R}^d , and two points are connected with some probability that varies (and usually decreases) with their distance. In particular, the special case in which the radius of connection is deterministically fixed at 1 was the model that initiated the study of this kind of random geometry, in the seminal paper of Gilbert [10]. The objects of interest in this model are the existence of an infinite connected component, as well as the behavior of the subset of \mathbb{R}^d that is at distance at most 1 from one of the vertices of the graph (the so-called “Poisson blob”). Continuum percolation is treated in detail in a book-length monograph by Meester and Roy [19], as well as in the more general work of Grimmett [12].

Most of the work done on random geometric graphs is concerned with either the behavior of a typical graph — the graph we are likely to see for a given r as n goes to infinity — or typical deviations from that behavior, i.e. central limit theorems. In this paper, we are concerned with the behavior of the model conditioned on a rare event. Specifically, we are concerned with the random geometric graph conditioned on having more edges than is expected (a formal description will follow). The large deviation regime of the upper tail of any subgraph count of the random geometric graph is not well studied, though some bounds are available: Janson [14] has studied concentration inequalities for U -statistics, a general class of statistics which includes the subgraph counts we are interested in. These upper bounds work in very general settings, but are not tight, even up to constants in the exponent. Large deviation principles have been proven for functionals of random point processes in which the contribution of any particular vertex is uniformly bounded [27], but no such bound is known for functionals with possibly large influence, such as the edge count of the graph.

As motivation for this detailed study, we consider the problem in a more familiar context: the “infamous upper tail” [15] of the triangle count T in the Erdős–Rényi random graph, $G(n, p)$. After many years of development of increasingly strong bounds, the first breakthrough was made by Kim and Vu [17] and Janson [16] independently, who proved that, for any $\delta > 0$ and whenever $p \gg (\log n)/n$,

$$\exp[-c(\delta)n^2p^2\log(1/p)] \leq \mathbb{P}[T > (1 + \delta)\mathbb{E}[T]] \leq \exp[-C(\delta)n^2p^2],$$

where $c(\cdot)$ and $C(\cdot)$ depend only on δ . Recently, there has been renewed interest in these type of tail estimates. In 2010, Chatterjee [5] and Demarco and Kahn [7] (in independent works) established the correct order

of the upper tail of triangles and other small cliques by adding the missing logarithmic term to the upper bound, without providing good control of the leading-order constants. The recent work of Chatterjee and Dembo [6] on nonlinear large deviations allowed Lubetzky and Zhao [18] to calculate both the order and the leading-order constant for the upper tail question in a certain regime of sparse Erdős–Rényi random graphs. Unfortunately, the technique is based on the analysis of polynomials in Bernoulli random variables, and thus ill-suited to the problem we are studying here.

In this work, we use the structure of the random geometric graph to evaluate the upper tail large deviation rate function. In addition, we provide a “structure theorem” to describe the graph-theoretical structure of the model conditioned on having too many edges. Specifically, we show that such a conditional model exhibits *localization*. Heuristically, this phenomenon can be described as a scenario in which a small number of vertices will contribute almost all the extra edges that we require the graph to exhibit, while the edge count of the “bulk” of the graph will remain largely unchanged, in some weak sense. Furthermore, we will show that the geometry of the localized region has the shape of a ball in the given norm (we will make these two statements more precise at the end of the next section). Outside of the aforementioned work of Lubetzky and Zhao [18] in the Erdős–Rényi model, which is only known for a limited range of p ’s, this work is the first (as far as the authors are aware) to establish that the large deviation regime of a subgraph count is (weakly) equivalent to planting a combinatorial structure in the usual, unconditional graph.

The fact that large deviation regimes may be dominated by configurations with a small number of very large contributions was known relatively early in the history of large deviation theory: a survey by Nagaev [21], summarizing a series of papers written in the Soviet Union in the 1960’s and 70’s, includes this observation for sums of i.i.d. random variables with stretched-exponential tails as a corollary. In our context, the natural combinatorial structure for creating many edges with a small number of edges is a “giant clique.” The clique number, the (typical) size of the largest clique of the random geometric graph, falls under the general class scan statistics, has been shown to focus on two values with high probability for certain values of the threshold parameter r (see [22], [20]); however, these works do not explore the large deviation regime. Our work uses techniques from large deviations, concentration inequalities, convex analysis, and geometric measure theory. A key component in the proof is a technique for proving localization that has previously appeared in [29] and [4].

2. MAIN RESULT

Let χ_n be a Poisson Point Process of intensity n on the d -dimensional unit torus $\mathbb{T}^d = [0, 1]^d$. Let $N := |\chi_n|$. Recall that N is a Poisson random variable with mean n , and conditional on N , χ_n is just a set of N points chosen

independently and uniformly at random. Let r_n be a positive sequence that decreases to 0 as $n \rightarrow \infty$, and $\|\cdot\|$ be some norm on \mathbb{R}^d that induces a translation-invariant metric on \mathbb{T}^d . We define the random geometric graph $G(\chi_n, r_n) := (V, E)$, where $V = \chi_n = \{v_1, \dots, v_N\}$, enumerated arbitrarily, and E is the set of unordered pairs $\{i, j\}$ such that $\|v_i - v_j\| \leq r_n$. Figure 1 shows a particular instance of $G(\chi_{150}, 0.1)$.

Letting $1_{i,j}$ be the indicator that there is an edge between v_i and v_j , we can calculate the expected value of $|E|$, the number of edges in the graph:

$$\begin{aligned} \mathbb{E}(|E|) &= \mathbb{E} \left(\sum_{1 \leq i < j \leq N} 1_{i,j} \right) = \mathbb{E} \left[\binom{N}{2} \mathbb{E}(1_{1,2} \mid N) \right] \\ &= \frac{n^2}{2} \mathbb{P}(\|v_1 - v_2\| \leq r_n). \end{aligned}$$

By the symmetry of the torus, $\mathbb{P}(\|v_1 - v_2\| \leq r_n)$ is simply νr_n^d , where ν is the volume of the unit ball of the given norm, r_n . Thus,

$$\mathbb{E}(|E|) = \frac{n^2 \nu r_n^d}{2} =: \mu_n.$$

For the rest of the article, we suppose the existence of a fixed constant $\delta^* > 0$ such that, for all sufficiently large n ,

$$(1) \quad n^{(\delta^*-2)/d} \leq r_n \leq n^{-\delta^*/d}$$

The lower bound ensures that the expected number of edges is polynomially large in n ; the upper bound excludes the possibility of r_n vanishing as the inverse of a polynomial in $\log n$ (for example), but is not much stronger than the initial requirement of r_n vanishing in as n grows. We define a parameter p as

$$(2) \quad p := \lim_{n \rightarrow \infty} \frac{\log \mu_n}{\log n},$$

implicitly assuming that the limit exists; we will continue to assume this for the rest of the paper. This ensures that $\mu_n = f(n)n^p$, where f varies more slowly than any polynomial or rational function in n . Notice that p depends only on the tail behavior of r , and is completely independent of n . Furthermore,

$$(3) \quad \delta^* \leq p \leq 2 - \delta^*,$$

thanks to (1).

The following theorem is the main result of the paper:

Theorem 1. *Let $G(\chi_n, r_n)$ be a random geometric graph model on the d -dimensional torus with respect to some norm $\|\cdot\|$ and a threshold parameter that satisfies $n^{(\delta^*-2)/d} \leq r_n \leq n^{-\delta^*/d}$ for some fixed $\delta^* > 0$. Let $\tau_n = \nu(r_n/2)^d$, where ν is the volume of the unit ball under the norm $\|\cdot\|$. (That is, τ_n is the volume of the ball of diameter r_n .) Fix $\delta > 0$ and $\varepsilon > 0$. Let*

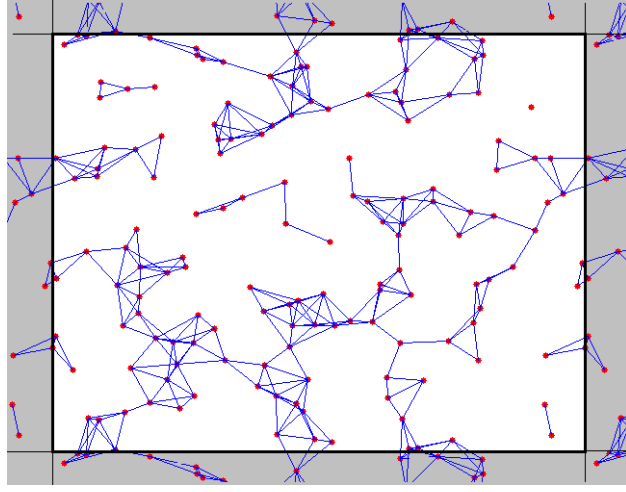


Figure 1. An instance of the random geometric graph $G(\chi_{150}, 0.1)$, with respect to the Euclidean norm. The graph has 148 vertices and 343 edges. The gray area is the white unit square translated, to show periodicity

F_n be the event that the following happen: (a) There is a ball A of diameter r_n such that any convex set $S \subseteq A$ with $\lambda(S) > (\varepsilon/16)\tau_n$ (where λ denotes Lebesgue measure) satisfies

$$\left| \frac{|\chi_n(S)|}{\sqrt{2\delta\mu_n}} - \frac{\lambda(S)}{\tau_n} \right| < \varepsilon,$$

and (b) for any convex set $S \subseteq A^c$ with $\lambda(S) > (\varepsilon/16)\tau_n$, such that S lies in some ball of diameter r_n ,

$$\frac{|\chi_n(S)|}{\sqrt{2\delta\mu_n}} < \varepsilon \frac{\lambda(S)}{\tau_n}.$$

Then the conditional probability of the event F_n given that $|E| \geq (1 + \delta)\mu_n$ tends to 1 as $n \rightarrow \infty$.

The convexity requirement is probably not optimal, but forces the set S to be sufficiently “nice” to preclude sets which are either sparse but of large measure (such as generalized Cantor sets) or have boundaries that take up a large amount of space. We also force S to be large to make sure it cannot pick up the lower order unevenness in the conditional process. Probing the Poisson Point Process with such sets is sufficient to show that graph will appear uniform, up to sets of measure equal to arbitrarily small multiples of τ_n .

As a consequence of Theorem 1, we prove that the upper tail of the edge count of random geometric graphs satisfies a large deviation principle. Recall that $(|E| - \mu_n)/\mu_n$ satisfies an *upper tail* large deviation principle with speed

$f(n)$ and rate function $I(x)$ if, for any closed set $F \subset (0, \infty)$,

$$\limsup_{n \rightarrow \infty} \frac{1}{f(n)} \log \mathbb{P} \left(\frac{|E| - \mu_n}{\mu_n} \in F \right) \leq - \inf_{x \in F} I(x),$$

and for any open set $G \subset [0, \infty)$,

$$\liminf_{n \rightarrow \infty} \frac{1}{f(n)} \log \mathbb{P} \left(\frac{|E| - \mu_n}{\mu_n} \in G \right) \geq - \inf_{x \in G} I(x).$$

Theorem 2. *Let $G(\chi_n, r_n)$ be a random geometric graph model on the d -dimensional torus, with the same assumptions as in Theorem 1. Define*

$$I(x) := \left(\frac{2-p}{2} \right) \sqrt{2x},$$

where p is defined as in (2). Then the random variable $(|E| - \mu_n)/\mu_n$ satisfies an upper tail large deviation principle with speed $\sqrt{\mu_n} \log n$ and rate function $I(x)$.

Note that this rate function, like all other claims in this paper, is only valid for events in which the number of edges exceeds its mean. The lower tail of the edge count is likely to satisfy Poisson-like statistics, and hence the speed of its large deviation probability is expected to be of order μ_n , not $\sqrt{\mu_n} \log n$.

Before we go on, we'd like to comment on the heuristic formulation we presented earlier. First, since the rate function of 2 is strictly increasing, we know that, conditional on $\{|E| > (1 + \delta)\mu_n\}$, the event

$$\{(1 + \delta')\mu_n > |E| > (1 + \delta)\mu_n\}$$

occurs with high probability for any $\delta' > \delta$ and n sufficiently large. If we take $S = A$ in the first stipulation of 1, we see that the ball A of diameter r_n includes at least $\delta\mu_n(1 - \varepsilon)$ edges of the random geometric graph. Since ε and $\delta' - \delta$ are arbitrarily close to zero, we recover our original claim that ‘almost all extra edges are between points in A ’. In terms of the geometry, we see that there are approximately $\sqrt{2\delta\mu_n}$ vertices that are more or less equidistributed inside A . Outside of this set, we would like to say that there are no other large cliques; unfortunately, 1 does not provide this result. Instead, we can only be sure that every other clique outside of the “exceptional” set A has $o(\sqrt{\mu_n})$ vertices. We conjecture that the size of the largest clique disjoint from A is of the same order as the size of the largest clique of an *unconditional* random geometric graph with the same values of n and r_n . A combination of this conjecture and elementary estimates on Poisson Point Processes should be sufficient to establish a (suitably formalized) version of the claim ‘the remaining portion of the graph remains unchanged.’

3. THE s -GRADED MODEL

Henceforth in the manuscript, we will suppress the subscript n and write χ, μ, τ and r instead of χ_n, μ_n, τ_n and r_n .

We now present an approximation of the random geometric model which allows us to replace the Poisson point process with a sequence of independent Poisson random variables. To do this, we first discretize space, and then produce a metric on the resulting “cells” that approximates the norm $\|\cdot\|$ on the unit torus. We call this the s -graded model.

Fix an integer s , and define

$$m := \lfloor s/r \rfloor,$$

so that

$$\frac{s}{r} - 1 \leq m \leq \frac{s}{r}.$$

Let $T = \{1, 2, \dots, m\}^d$. Pick $I = (i_1, i_2, \dots, i_d) \in T$, and define

$$A_I = \left[\frac{i_1 - 1}{m}, \frac{i_1}{m} \right] \times \dots \times \left[\frac{i_d - 1}{m}, \frac{i_d}{m} \right].$$

The A_I 's partition the unit torus into m^d cubes (ignoring sets of measure 0), each of volume $1/m^d$, and therefore, $X_I = |\chi(A_I)|$ is a Poisson random variable of mean

$$\mathcal{D} := \frac{n}{m^d}.$$

As a function of n , m^d grows as n^{2-p} , up to a possible multiplicative factor that varies more slowly than any polynomial or rational function in n (recall that s is fixed); this implies that \mathcal{D} grows as n^{p-1} . Note that s is exactly the minimal number of cells that must be traversed in any one axial direction to ensure that we exceed the distance r . This parameter takes the place of r in the s -graded model.

We now define a metric on T , induced by the norm on torus:

$$(4) \quad d(I, J) = \inf_{x \in A_I^\circ, y \in A_J^\circ} \lceil m\|x - y\| \rceil$$

where the circles indicate the interiors of the sets. Note that the distance is always an integer. Moreover, $d(I, J) = z$ if z is the smallest integer such that some point in A_I° and some point in A_J° are less than z away, measured in units of $1/m$, the side length of the cubes. We force the points to be in the interior to prevent “trivialities”, such as two adjacent cells being distance 0, since they share a boundary.

We are now ready to define the s -graded random geometric graph. Let $G_s(\chi, r) = (V, E_s)$ have the same vertex set as the original graph. For each vertex v , let I_v be the index in T such that $v \in A_{I_v}$; there is ambiguity on the boundary of the A_I 's, but that set has Lebesgue measure 0, and therefore it has no vertices of χ , almost surely. We say $(v, w) \in E_s$ whenever $d(I_v, I_w) \leq s$. Essentially, the s -graded model allows every point to wander inside a cubical “cage” of side-length $1/m$, and connects any two points that might be connected after we allow this mobility. In this framework, it is clear that E_s becomes smaller as s decreases. In fact, for sufficiently large s , E_s is identical to E ; unfortunately, this s will depend on n . Due to

uniformity issues, we will instead fix s and later show that, even when s is finite but large, the approximation is good.

The major benefit of the s -graded model is that its edge count is very simple to express in terms of X_I , the number of points in each A_I :

$$\begin{aligned} |E_s| &= \sum_{I \in T} \left[\binom{X_I}{2} + \frac{1}{2} \sum_{J: 0 < d(I, J) \leq s} X_I X_J \right] \\ &= \frac{1}{2} \sum_{I \in T} X_I \left[\left(\sum_{J: d(I, J) \leq s} X_J \right) - 1 \right]. \end{aligned}$$

This random variable is defined in terms of i.i.d. random variables, which eases the analysis greatly. Furthermore, the ambient geometry of the torus is encoded completely by the metric d . Finally, each X_I only appears in finitely many terms in this expression. The “finite range” nature of the representation will play a major role in the proof presented.

We quantify this fact as follows: for any $I \in T$, let $N_I := \{J : d(I, J) \leq s\}$. Thanks to translation invariance of d , the cardinality of this set is independent of the choice of I . Using this parameter, we can compute the expected number of edges in the s -graded random geometric graph easily:

$$\begin{aligned} (5) \quad \tilde{\mu}_s &:= \mathbb{E}(|E_s|) \\ &= \sum_{I \in T} \mathbb{E} \left[\binom{X_I}{2} \right] + \frac{1}{2} \sum_{J: 0 < d(I, J) \leq s} \mathbb{E}(X_I) \mathbb{E}(X_J) \\ &= \frac{|N_I| m^d \mathcal{D}^2}{2} = \frac{|N_I| n^2}{2m^d}. \end{aligned}$$

As before, we are interested in conditioning the s -graded model on the event $\{|E_s| > (1 + \tilde{\delta})\mu_s\}$. Again, the appropriate geometric notion is that of the largest set of diameter s . We call a set of indices a *maximal clique set* if it is a subset of T with diameter s that achieves the maximal cardinality of all such sets. We define

$$(6) \quad \tilde{\tau}_s := \max\{|\mathcal{J}| : \mathcal{J} \subset T, \text{diam}(\mathcal{J}) \leq s\},$$

i.e. τ_s is the cardinality of a maximal clique set.

We can now state the equivalent to Theorem 1 for the s -graded model:

Theorem 3. *Fix an integer s and let $G_s(\chi, r)$ be an s -graded random geometric graph, with $n^{(2-\delta^*)/d} \leq r \leq n^{-\delta^*/d}$ for some fixed $\delta^* > 0$. Fix $\tilde{\delta} > 0$ and $\tilde{\varepsilon} > 0$. Let F be the event that the following happen: (a) There is a maximal clique set \mathfrak{P} such that for all $I \in \mathfrak{P}$,*

$$\left| \frac{\tilde{\tau}_s X_I}{(2\tilde{\delta}\tilde{\mu}_s)^{1/2}} - 1 \right| < \tilde{\varepsilon},$$

and (b) for all $J \in \mathfrak{P}^c$,

$$\frac{\tilde{\tau}_s X_J}{(2\tilde{\delta}\tilde{\mu}_s)^{1/2}} \leq \tilde{\varepsilon}.$$

Then the conditional probability of F given $|E_s| \geq (1 + \tilde{\delta})\tilde{\mu}_s$ tends to 1 as $n \rightarrow \infty$.

Essentially, Theorem 3 produces a maximal clique set, and whose entries sum up to $(2\tilde{\delta}\tilde{\mu}_s)^{1/2}$, up to lower order corrections. This creates all the ‘extra edges’ in the graph. As in Theorem 1, this set will contain almost all the edges not expected to be in unconditional s -graded model, and the rest of graph will look unaltered, at least in the scale of $\sqrt{\tilde{\mu}_s}$. This theorem will also be instrumental in proving 2; that proof will appear in Section 7.

4. OUTLINE OF THE PROOF

Before embarking on a proper proof, we sketch the main ideas required. The first step is to show that the s -graded model is, in fact, a good approximation for the random geometric graph. To do so, we first show that we can approximate any convex subset S of a ball of diameter r from both the inside and the outside by a union of A_I ’s. Next, we use the classical isodiametric inequality to show that the A_I ’s associated with a maximal clique set approximate a ball of diameter r , in the sense of the Hausdorff metric. Armed with these facts, showing that Theorem 3 implies Theorem 1 is a matter of careful “epsilonotics.”

We then turn to directly analyzing the s -graded model, conditioned on the event

$$L := \{|E_s| \geq (1 + \tilde{\delta})\tilde{\mu}_s\}.$$

For notational convenience, let $q = \sqrt{2\tilde{\delta}\tilde{\mu}_s}$, $w = \tilde{\tau}_s\mathcal{D}$. Define

$$\mathfrak{I} = \{I \in T : X_I > \max\{\mathcal{D}n^a, n^a\}\},$$

where a is a small positive number. This is the set of “large” indices, in the sense that their associated X_I ’s exceed their expected value by a fixed polynomial factor in n . Furthermore, define

$$Y_I := X_I (\log(X_I/\mathcal{D}) - 1) + \mathcal{D},$$

and

$$Q(\mathfrak{I}) := \frac{2}{q^2} \left(\sum_{I \in \mathfrak{I}} \binom{X_I}{2} + \frac{1}{2} \sum_{J \in N_I \cap \mathfrak{I}} X_I X_J \right).$$

The former quantity is an appropriately chosen convex function of the X_I ’s, while the latter is a scaled version of the number of edges with both endpoints in the A_I ’s associated with \mathfrak{I} . Consider the event

$$\left\{ Q(\mathfrak{I}) > 1 - \frac{\xi}{\log n} \right\} \cap \left\{ \frac{1}{q} \sum_{I \in \mathfrak{I}} Y_I \leq \log(q/w) - 1 + \xi \right\},$$

where ξ is an arbitrary positive constant. By comparing a lower bound on the probability of L to an upper bound on the probability of the *complement* of the above event, we can show that this event occurs with high probability in the s -graded model conditional on L .

We now have a set of indices \mathfrak{I} which satisfies both a quadratic lower bound and a convex upper bound with high probability. From here on, the analysis is completely deterministic, characterizing configurations that satisfy the simultaneous inequalities above. We wish to look at the largest elements of \mathfrak{I} . Specifically, we wish to take the smallest set that includes roughly q vertices of the Poisson point process. To make this precise, define

$$V(A) := \frac{1}{q} \sum_{I \in A} X_I,$$

for any subset of indices A . Now, order the elements of \mathfrak{I} by size, so that X_1 is the largest, X_2 is the second largest, etc. Letting $\mathfrak{T}_k = \{X_1, X_2, \dots, X_k\}$, define

$$\mathfrak{T} := \text{The first } \mathfrak{T}_k \text{ such that } V(\mathfrak{T}_k) > 1 - \frac{2\xi}{\log n}.$$

Careful use of minimality and Jensen's inequality can establish that

$$V(\mathfrak{T}) \leq 1 + \phi(\mathfrak{T}),$$

where $\phi(\mathfrak{T})$ is bounded above by $\xi/3$. More importantly, moving from \mathfrak{I} to \mathfrak{T} does not force us to discard too many edges; formally,

$$Q(\mathfrak{T}) \geq 1 - \psi(\mathfrak{T}),$$

with $\psi(\mathfrak{T})$ bounded above by ξ .

We now define

$$\mathfrak{P} := \left\{ I \in \mathfrak{T} : X_I > \frac{\xi^{1/4} q}{\tilde{\tau}_s} \right\}.$$

If \mathfrak{I} was the set of large indices, \mathfrak{P} is the set of very large indices — that is, those whose associated X_I 's are commensurate with q . We can show that this set cannot have diameter that strictly exceeds s without contradicting either the lower bound $Q(\mathfrak{T})$ or the upper bound on $V(\mathfrak{T})$. Furthermore, technical estimates ensure that $|\mathfrak{P}| \geq \tilde{\tau}_s$. As a set of diameter at most s and cardinality at least $\tilde{\tau}_s$, \mathfrak{P} must be a maximal clique set, by definition. Moreover, a quantitative version of Jensen's inequality allows us to claim that all the X_I 's associated with this set are roughly equal. Thus, \mathfrak{P} is the maximal clique set described in Theorem 3. Finally, we can show that its complement is made up on X_I 's that vanish in comparison to q . This completes the proof.

In the last section, we prove the large deviation principle. We use the first stipulation of 1 and the s -graded model to compute the upper bound. The lower bound is derived directly from the Poisson Point Process.

5. THE s -GRADED MODEL APPROXIMATES THE RANDOM GEOMETRIC GRAPH

Before proving Theorem 3, we show that it implies Theorem 1. To do so, we first define three operations to go between subsets of T , the natural

objects in the s -graded model, and subsets of $[0, 1]^d$. For $\mathfrak{B} \subset T$, let

$$\mathfrak{U}(\mathfrak{B}) := \bigcup_{I \in \mathfrak{B}} A_I$$

the set associated with the index set \mathfrak{A} . In the other direction, we cannot be as exact; for $K \subset [0, 1]^d$, let $\mathfrak{R}(K)$ and $\mathfrak{D}(K)$ be the outer and inner hulls of K , defined as the minimal and maximal subsets of T such that

$$\mathfrak{U}(\mathfrak{R}(K)) \subset K \subset \mathfrak{U}(\mathfrak{D}(K))$$

With these operators in place, we show that the s -graded parameters are good approximations of their respective properties in the usual random geometric graph:

Lemma 4. *Let $G(\chi, r) = (V, E)$ be the usual random geometric graph, and $G_s(\chi, r) = (V, E_s)$ be the s -graded model. As before, $\mu = \mathbb{E}(|E|)$, $\tilde{\mu}_s = \mathbb{E}(|E_s|)$ and $\tau = \nu(r/2)^d$. Let $\tilde{\tau}_s$ be the cardinality of a maximal clique set (as defined in (6)). Then $E \subseteq E_s$, and there exist constants C and s_0 , depending only on the dimension and the chosen norm of the torus, such that if $s \geq s_0$ then*

$$\mu \leq \tilde{\mu}_s \leq \mu \left(1 + \frac{C}{s}\right)$$

and

$$m^d \tau \leq \tilde{\tau}_s \leq m^d \tau \left(1 + \frac{C}{s}\right)$$

Furthermore, both $|N_I|$ and $\tilde{\tau}_s$ are uniformly bounded in n .

Proof. Pick an arbitrary I and consider $\mathfrak{U}(N_I)$. By definition of d and s , this set includes a ball of radius r around *any* point in A_I . Therefore, any pair $(v, w) \in E$ must also be in E_s , giving the first stipulation. Since this inclusion holds for any configuration of the underlying Poisson Point process, this also gives $\mu \leq \tilde{\mu}_s$.

Now, let ρ be the diameter of the unit cube under the norm $\|\cdot\|$. Then, for any $x \in A_I$ and $y \in \mathfrak{U}(N_I)$,

$$\|x - y\| \leq \frac{rs}{s - r} + \frac{2\rho}{m}$$

Therefore, $\mathfrak{U}(N_I)$ is contained in a ball of radius $r + 3\rho/m$ around any point in A_I , which implies that

$$|N_I| = m^d \lambda(\mathfrak{U}(N_I)) \leq \nu m^d r^d \left(\frac{s}{s - r} + \frac{3\rho}{rm}\right)^d \leq \nu m^d r^d \left(1 + \frac{3r + 6d\rho}{s - r}\right)$$

where the final inequality follows because $(1 + x)^d \leq 1 + (2d)x$ for all sufficiently small x , and the fact that $rm \geq s - r$ by definition of m . Substituting this into the definition of $\tilde{\mu}_s$ produces the desired inequality on $\tilde{\mu}_s$ (assuming, without loss, that $r \leq \rho$).

To show the lower bound on $\tilde{\tau}_s$, let B be a ball of radius $r/2$. $\mathfrak{U}(B)$ will have diameter at most s , by definition of the metric, and its measure will

be at least τ . Converting this to cardinality gives the desired lower bound. For the upper bound, let \mathfrak{W} be a set of indices such that

$$\lambda(\mathfrak{U}(\mathfrak{W})) \geq \tau \left(1 + \frac{C}{s}\right)$$

Applying the isodiametric inequality for finite dimensional normed spaces [3, p. 93] and choosing C and s_0 sufficiently large gives

$$\text{diam}(\mathfrak{U}(\mathfrak{W})) \geq r \left(1 + \frac{C}{s}\right)^{1/d} \geq r + \frac{4\rho}{m}$$

This implies that the diameter of \mathfrak{W} is at least $s + 1$ in the d metric, and therefore the set cannot be a maximal clique set. Translating the measure bound to a cardinality bound on the size \mathfrak{W} gives the desired upper bound on \mathfrak{W} .

Finally, $m^d \leq s^d/r^d$, giving the uniform upper bound on $|N_I|$. Since $\tilde{\tau}_s \leq |N_I|$, the final claim of the lemma follows. \square

As a simple corollary to the lemma, we see that the hypothesis of Theorem 1 implies the hypothesis of Theorem 3, as long as $\tilde{\delta} < \delta$ and s is sufficiently large:

Corollary 5. *For any $\tilde{\delta} < \delta$, there exists s_0 depending only on δ , the chosen norm, and the dimension, such that whenever $s \geq s_0$, $|E| > (1 + \delta)\mu$ implies $|E_s| > (1 + \tilde{\delta})\tilde{\mu}_s$.*

Proof. Pick s large enough to ensure $(1 + \tilde{\delta})(1 + C/s) \leq 1 + \delta$. Lemma 4 guarantees that $|E_s| \geq |E|$ and

$$|E| \geq (1 + \delta)\mu \geq (1 + \tilde{\delta})(1 + C/s)\mu \geq (1 + \tilde{\delta})\tilde{\mu}_s$$

This proves the corollary. \square

Next, we get some quantitative bounds on the geometric approximation of certain subsets of $[0, 1]^d$ by the A_I 's. Let S be a convex subset of a ball of diameter r . Assume that

$$z(S) := \frac{\lambda(S)}{\tau} \geq \varepsilon/16.$$

Recall that τ is the volume of a ball of diameter r . We rescale Lebesgue measure to bypass factors of τ throughout this section. The first goal is to show that the inner and outer hulls of S approximate S uniformly in the choice of S , in the sense of the measure of the sets. Formally, we wish to prove the following proposition:

Proposition 6. *Fix $\varepsilon > 0$, and let S be a convex subset of a ball of radius r with $z(S) \geq \varepsilon/16$. Then, there exists an s_0 that depends only on ε , the dimension, and the choice of norm, such that whenever $s > s_0$, the inner hull $\mathfrak{U}(S)$ satisfies*

$$z(S) - \frac{\varepsilon^2}{64} \leq z(\mathfrak{U}(\mathfrak{R}(S))) \leq z(S)$$

Similarly, the corresponding inequality

$$z(S) \leq z(\mathfrak{U}(\mathfrak{D}(S))) \leq z(S) + \frac{\varepsilon^2}{64}$$

hold for the outer hull of S .

The essential element of this proposition is that s_0 is completely independent of the precise choice of S , as long as S satisfies a diameter upper bound and measure lower bound. For a single choice of S , the continuity of Lebesgue measure is enough to establish these inequalities; however, the choice of s_0 given by this construction would depend on the geometry of S , which is insufficient for our purposes.

Instead of proving this statement directly, we consider this minor modification of the proposition:

Lemma 7. *Let S be a convex set as in 6. Define*

$$(\partial S)_l := \{y : \|\partial S - y\| \leq l\}.$$

Fix $\varepsilon > 0$. Then, there exists an s_0 , depending on ε , the dimension, and the norm, such that $s > s_0$ implies

$$z((\partial S)_{\rho/m}) \leq \frac{\varepsilon^2}{64}$$

This lemma implies 6, since

$$S \subset \mathfrak{U}(\mathfrak{R}(S)) \cup (\partial S)_{\rho/m}$$

and

$$\mathfrak{U}(\mathfrak{D}(S)) \subset S \cup (\partial S)_{\rho/m}.$$

Evaluating $z(\cdot)$ of both sides and using subadditivity gives the two nontrivial bounds required in 6; the other two follow by inclusion.

Proof of Lemma 7. Heuristically, the volume of $(\partial S)_{\rho/m}$ should be commensurate with the product of ρ/m with the surface area of S . Since S has diameter at most r and is a convex set (and therefore its boundary cannot be too convoluted), this surface area ought to be bounded above by Cr^{d-1} , and therefore $z[(\partial S)_{\rho/m}]$ will be small. To formalize this loose heuristic, we turn to the tools of geometric measure theory. Although this approach is standard in that field, we include a detailed proof for completeness. The analysis is slightly technical, and breaks down to three parts: First, we state the coarea formula for Lipschitz functions, defining all relevant terms. Next, we establish that the function $f(x) = \|\partial S - x\|$ is Lipschitz, and use the coarea formula to express the measure of $(\partial S)_{\rho/m}$ as an integral of the $(d-1)$ dimensional Hausdorff measures of well-behaved sets. Finally, we show that, thanks to convexity of S , the measures in the integrand can be uniformly bounded by Cr^{d-1} ; a computation completes the proof.

Consider a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ that is Lipschitz with respect to the Euclidean distance, and a Borel set A . The Euclidean coarea formula [9, pg. 248] states that, with the functions as above,

$$\int_A \|Df(x)\|_2 dx = \int_{-\infty}^{\infty} H^{d-1}(A \cap f^{-1}(y)) dy,$$

where $\|\cdot\|_2$ is the Euclidean norm, and H^{d-1} is the Hausdorff measure on the surface (effectively the surface area) and Df is the gradient of f , which exists since f is almost everywhere differentiable. To formally define this, pick a set A and let $U_\delta(A)$ be the set of all coverings $\{U_i\}$ of A , where U_i have diameter at most δ . By definition, the Hausdorff measure of A is

$$H^{d-1}(A) = C_{d-1} \lim_{\delta \downarrow 0} \inf_{\{U_i\} \in U_\delta(A)} \sum [\text{diam}(U_i)]^{d-1},$$

where C_{d-1} is some constant depending on dimension, used to normalize the measure appropriately to be compatible with Lebesgue measure. This limit is well defined [9, pg. 170], but may be 0 or infinity for a general A ; in fact, it is infinite for any set with positive d -dimensional Lebesgue measure.

Now, consider the function $f(x) = \|\partial S - x\|$, where $\|B - x\|$ is shorthand for $\inf_{b \in B} \|b - x\|$ for any set B . First, we recall the classical fact that all norms are equivalent in finite dimensional space, i.e. there exists two positive constants c and C such that, for all x and y ,

$$c\|x - y\| \leq \|x - y\|_2 \leq C\|x - y\|.$$

To see that f is Lipschitz with respect to the Euclidean norm, pick two points x and y , and let $a \in \partial S$ be the point such that $f(x) = \|x - a\|$ (this point exists because ∂S is closed). Then

$$f(y) - f(x) \leq \|y - a\| - \|x - a\| \leq \|x - y\| \leq C\|x - y\|_2.$$

Similarly, $f(x) - f(y) \leq C\|x - y\|_2$. Thus, f is differentiable almost everywhere. Pick an x where the function is differentiable, and let a be as before. Then, for any $t \in (0, 1)$,

$$f(x + t(a - x)) \leq f(x) - t\|a - x\|,$$

by the properties of norms. Subtracting $f(x)$ from both sides, dividing by t and letting $t \rightarrow 0$, we get

$$\langle a - x, Df \rangle \leq -\|a - x\|,$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product (or in this case, the directional derivative). Applying the Cauchy-Schwarz inequality, we conclude that

$$\|a - x\| \leq \|a - x\|_2 \|Df\|_2,$$

which implies, by the equivalence of norms, that

$$\|Df\|_2 \geq \frac{\|a - x\|}{\|a - x\|_2} \geq c.$$

Substituting this into the coarea formula, we see that

$$c\lambda(A) \leq \int_{-\infty}^{\infty} H^{d-1}(A \cap f^{-1}(y)) dy,$$

for any choice of A .

Letting $A = \{x : f(x) \leq \rho/m\}$, we deduce that

$$\lambda[(\partial S)_{\rho/m}] \leq C \int_0^{\rho/m} H^{d-1}(\{y : \|\partial S - y\| = z\}) dz,$$

where C is some (possibly different) universal constant. Note that, for sufficiently small z and T convex, the set $\{y : \|\partial T - y\| = z\}$ has two parts: one inside T and the other outside of it. Luckily, the external one is the boundary of T_z , which is convex as it is the affine sum of T and a ball of radius z . The internal one is boundary of the set

$$T^{(z)} := \{x : x \in T, \|\partial T - x\| \geq z\}.$$

This set is also convex; if it weren't, we could find $x, y \in T^{(z)}$ such that $w = tx + (1-t)y \notin T^{(z)}$ for some $t \in (0, 1)$. Let v be the minimal length vector such that $w + v \in \partial T$. Then, since $\|v\| < z$ by definition, we can find an ε sufficiently small such that $w + (1+\varepsilon)v \notin T$, while $x + (1+\varepsilon)v$, $y + (1+\varepsilon)v$ are both in T , contradicting convexity of T . Therefore, controlling the “fattening” of ∂S reduces to controlling the boundaries of convex sets.

Suppose T is a convex subset of a Euclidean ball B centered around some x in the interior of T . Define $P_x : \partial B \rightarrow \partial T$ by defining $P_x(y)$ to be the intersection of the ray from x to y with ∂T . Since T is convex, this is well defined; suppose, by way of contradiction, that some ray emanating from x hits a pair of distinct points $b_1, b_2 \in \partial T$, in this order. Since x is an interior point of T , there is some open set O that includes x entirely inside T . Now, define the set

$$\{t : t = \lambda y + (1-\lambda)b_2, y \in O, \lambda \in (0, 1)\}.$$

This set is open, and contains b_1 . By convexity of T , it is completely inside T , and thus b_1 is not a boundary point, contradicting the assumption. This also implies that the map P_x is bijective.

We now use the map P_x to control the $(d-1)$ dimensional Hausdorff measure of the boundary of T . To do so, we consider $\{U_i\} \in \mathcal{U}_\delta(\partial B)$ - i.e. a particular covering of the boundary of the Euclidean ball B . Without loss of generality, we may assume that all the U_i 's are subsets of ∂B ; otherwise, we simply intersect the U_i 's with ∂B , possibly decreasing the diameter. Define $\{U'_i\}$ to be the image of the U_i 's under P_x . The $\{U'_i\}$'s cover ∂T , since P_x is bijective, and the diameter of U'_i is no greater than the diameter of U_i . Taking a minimizing sequence of coverings of ∂B , we conclude that

$$H^{d-1}(\partial T) \leq H^{d-1}(\partial B).$$

Note that, formally, it is possible that either of these measures is infinite. However, the Hausdorff measure of the Euclidean ball is well known to be finite [9, pg. 171], meaning $H^{d-1}(\partial T)$ is finite.

We now apply these bounds to S , the convex subset of a ball of radius r . By equivalence of norms, S is also a subset of a Euclidean ball of radius Cr centered at some point in its interior, for some C ; in fact, for all sufficiently small z , both S_z and $S^{(z)}$ are also subsets of this ball. Therefore, using the above inequality,

$$H^{d-1}(\{y : \|\partial S - y\| = z\}) = H^{d-1}(\partial S_z) + H^{d-1}(\partial S^{(z)}) \leq CH^{d-1}(\partial B).$$

Since B is a ball of radius Cr , there exists some constant that depends only on the norm and the dimension such that

$$H^{d-1}(\partial B) = Cr^{d-1},$$

for some possibly different C . Plugging in these estimates into the Euclidean coarea formula, we get

$$\lambda((\partial S)_{\rho/m}) \leq C' \int_0^{\rho/m} r^{d-1} dz \leq \frac{C''}{m^d},$$

where C'' and C'' are possibly different constants depending only on the dimension and norm. Since $m > (s/r) - 1$ and $\tau = \nu(r/2)^d$, we conclude that, for sufficiently large s depending only on d , the norm, and ε ,

$$z((\partial S)_{\rho/m}) \leq \frac{C''}{\tau m^d} \leq \frac{C'''}{(s-r)^d} \leq \frac{\varepsilon^2}{64},$$

as required. \square

As a final preliminary, we prove a lemma about the geometry of maximal clique sets:

Lemma 8. *Let \mathfrak{P} be a maximal clique set. Then, for any $\varepsilon > 0$, $\mathfrak{U}(\mathfrak{P})$ contains a ball of diameter $(1 - \varepsilon)r$, for some sufficiently large s .*

Proof. To prove this lemma, we go through the abstract framework of Hausdorff convergence of subsets of a metric space. Consider an abstract metric space X imbued with metric ι , and, for any $S \subset X$, define the l -fattening of S as before, using the metric ι to measure distance. For any two $A, B \subset X$, the Hausdorff distance is defined as

$$\iota_H(A, B) := \inf\{l : A \subset (B)_l, B \subset (A)_l\}.$$

If X is compact in the topology defined by ι , the space of closed subsets of X makes a compact space with respect to this metric [26, page 294].

Let A be a cube of side $3r$ in the d -dimensional torus. Note that $\mathfrak{U}(\mathfrak{R}(A))$ includes a cube of side $(3 - 2/s)r$. If $s > 2$, such a set will also include a ball of diameter r , and therefore at least one maximal clique set. For each $s > 2$, let \mathfrak{P}_s be some maximal clique set that is a subset of $\mathfrak{R}(A)$.

Let $W_s = \mathfrak{U}(\mathfrak{P}_s)$. The diameter of W_s is bounded above by $r(1 + C/s)$, and its measure is bounded below by τ from Lemma 4. Let \tilde{A} and \tilde{W}_s be the same sets, with distances scaled by $1/r$. Then \tilde{W}_s are subsets of a cube of side-length ≤ 3 . Their diameter cannot exceed $(1 + C/s)$ and their Lebesgue measure is at least $\nu/2^d$, where the final bound follows from the definition of τ . Note that, after rescaling, every \tilde{W}_s is a subset of the *same* metric space \tilde{A} , which is independent of n . Thus, any convergence in the Hausdorff distance associated with \tilde{A} is automatically uniform in n . Since \tilde{A} is a compact metric space and \tilde{W}_s is a sequence of closed sets, the compactness result above guarantees that some subsequence \tilde{W}_{s_k} converges to a limit \tilde{W} in the Hausdorff metric. Passing through to the limit, we see that \tilde{W} must have measure at least $\nu/2^d$, and its diameter cannot exceed 1. However, it is also a subset of a cube of side-length ≤ 3 , and therefore we can embed \tilde{W} into the finite dimensional normed space $(\mathbb{R}^d, \|\cdot\|)$ isometrically. Quoting the isodiametric inequality again [3, pg 94],

$$\lambda(\tilde{W}) \leq \nu \left(\frac{\text{diam} \tilde{W}}{2} \right)^d.$$

Thus, \tilde{W} satisfies equality for the isodiametric inequality. Luckily, the isodiametric inequality also characterizes all sets that achieve equality as balls in the given norm.

Fix $\varepsilon > 0$. For all sufficiently large k , the Hausdorff convergence of \tilde{W}_{s_k} guarantees that, for some $B(x)$, a ball of diameter 1 centered at x , we have

$$B(x) \subset (\tilde{W}_{s_k})_{\varepsilon/2}.$$

Now, consider the set $B(x) \setminus \tilde{W}_{s_k}$. The distance between x and this set must exceed $1/2 - \varepsilon/2$; otherwise, $B(x)$ would not be inside the $\varepsilon/2$ -fattening of \tilde{W}_{s_k} . Thus, a ball of diameter $1 - \varepsilon$ must be inside \tilde{W}_{s_k} . Scaling by r completes the proof of Lemma 8. \square

Proof of Theorem 1, assuming Theorem 3. Fix $\delta > 0$ and $\varepsilon > 0$, and assume that $|E| > (1 + \delta)\mu$. Let $\tilde{\delta} = \delta(1 - \varepsilon/16)$. By 5, $|E_s| > (1 + \tilde{\delta})\tilde{\mu}_s$ occurs for some sufficiently large s . Let $\tilde{\varepsilon} = \varepsilon/8$. Assume that the event F described in Theorem 3 happens. For ease of notation, let $q = (2\tilde{\delta}\tilde{\mu}_s)^{1/2}$. Then there is a maximal clique set $\mathfrak{P} \subseteq T$ such that for all $I \in \mathfrak{P}$,

$$\left| \frac{\tilde{\tau}_s X_I}{q} - 1 \right| < \tilde{\varepsilon}.$$

Now, let A be some ball of diameter $(1 - \varepsilon/128)r$ contained in $\mathfrak{U}(\mathfrak{P})$ - for sufficiently large s , such a set exists by Lemma 8. Pick S to be a convex subset of A with $z(S) \geq \varepsilon/16$. We now wish to show that

$$(z(S) - 15\varepsilon/16)\sqrt{2\delta\mu} \leq |\chi(S)| \leq (z(S) + 15\varepsilon/16)\sqrt{2\delta\mu}.$$

This would be sufficient to prove the first stipulation of Theorem 1 if A were a ball of radius r ; because of the error in the radius, we must get slightly

better bounds and strengthen them later. To get an upper bound, we look at the number of vertices in the outer hull of S ; the corresponding lower bound will arrive via a bound on the inner hull.

First, we want to get an upper bound on $|\mathfrak{D}(S)|$. By the upper bound on the Lebesgue measure of $\mathfrak{U}(\mathfrak{D}(S))$ from 6, we conclude that

$$\begin{aligned} |\mathfrak{D}(S)| &\leq (m^d \tau) \left(z(S) + \frac{\varepsilon^2}{64} \right) \\ &\leq z(S) \tilde{\tau}_s \left(1 + \frac{\varepsilon^2}{64 z(S)} \right) \\ &\leq z(S) \tilde{\tau}_s \left(1 + \frac{\varepsilon}{4} \right), \end{aligned}$$

where the penultimate inequality follows from Lemma 4, and the final one from the assumed lower bound on $z(S)$. If we allow every X_I in $\mathfrak{D}(S)$ to take on the maximal value allowed by Theorem 3, we deduce the inequality

$$\begin{aligned} |\chi(S)| &\leq |\chi(\mathfrak{U}(\mathfrak{D}(S)))| \\ &\leq z(S) \tilde{\tau}_s \left(1 + \frac{\varepsilon}{4} \right) \left[\left(\frac{q}{\tilde{\tau}_s} \right) (1 + \tilde{\varepsilon}) \right] \\ &\leq z(S) q \left(1 + \frac{\varepsilon}{2} \right). \end{aligned}$$

This is nearly the desired upper bound; we simply need to replace q with $\sqrt{2\delta\mu}$. By appealing to 4, we see that

$$q = (2\tilde{\delta}\tilde{\mu}_s)^{1/2} \leq \sqrt{2\delta\mu} \left(1 + \frac{C}{s} \right)^{1/2}.$$

Increasing s sufficiently to ensure that

$$\left(1 + \frac{C}{s} \right)^{1/2} \leq \left(1 + \frac{\varepsilon}{4} \right)$$

and substituting into the earlier inequality, we produce the desired upper bound on $|\chi(S)|$.

The lower bound follows similarly. First, we get a lower bound on $|\mathfrak{R}(S)|$:

$$|\mathfrak{R}(S)| \geq \left(z(S) - \frac{\varepsilon^2}{64} \right) (m^d \tau) \geq z(S) \tilde{\tau}_s \left(1 + \frac{C}{s} \right)^{-1} \left(1 - \frac{\varepsilon}{4} \right)$$

where we appeal to 6 for the lower bound on the Lebesgue measure of $\mathfrak{U}(\mathfrak{R}(S))$. Since $S \subset A$, $\mathfrak{R}(S) \subset \mathfrak{P}$. If every X_I is bounded below by the minimal value in \mathfrak{P} , we get the following lower bound for $|\chi(S)|$:

$$|\chi(S)| \geq |\chi(\mathfrak{U}(\mathfrak{R}(S)))| \geq z(S) \tilde{\tau}_s \left(1 + \frac{C}{s} \right)^{-1} \left(1 - \frac{\varepsilon}{4} \right) \left[\left(\frac{q}{\tilde{\tau}_s} \right) (1 - \tilde{\varepsilon}) \right].$$

The lower bound on q

$$q = (2\tilde{\delta}\tilde{\mu}_s)^{1/2} \geq \left(\frac{2\delta\mu}{1 - \varepsilon/16} \right)^{1/2}$$

follows from the Lemma 4 bound $\tilde{\mu}_s \geq \mu$ and the definition of $\tilde{\delta}$. Substituting this in and increasing s sufficiently gives the desired lower bound.

As mentioned above, this nearly completes the first stipulation of Theorem 1; the only difference is that the ball A is of diameter $(1 - \varepsilon/128)r$ instead of r . Let A' be an arbitrary ball of diameter r containing A , and pick a convex set S' in A . By definition, $z(A'/A) \leq \varepsilon/32$. Defining $S = S' \cap A$, all the previous assertions follow. Furthermore, the lower bound

$$|\chi(S')| \geq |\chi(S)| \geq (z(S) - 15\varepsilon/6)\sqrt{2\delta\mu}$$

follows from before. Finally, $z(S) \geq z(S') - \varepsilon/32$, as $z(S'/S) \leq z(A'/A)$, which implies the lower bound

$$|\chi(S')| \geq (z(S') - \varepsilon)\sqrt{2\delta\mu}$$

For the upper bound, we note that

$$\begin{aligned} |\chi(S')| &\leq |\chi(S)| + |\chi(\mathfrak{D}(S'/S))| \\ &\leq (z(S) + 15\varepsilon/16) \left(\sqrt{2\delta\mu} \right) + |\chi(\mathfrak{D}(S'/S))|. \end{aligned}$$

Again, assuming all elements of $\mathfrak{D}(S'/S)$ intersect the indices with largest density, we can conclude that

$$|\chi(\mathfrak{D}(S'/S))| \leq z(S'/S)\tilde{\tau}_s \left(1 + \frac{\varepsilon}{4} \right) \left(\frac{q}{\tilde{\tau}_s} \right) (1 + \tilde{\varepsilon}) \leq \left(\frac{\varepsilon}{16} \right) (\sqrt{2\delta\mu})$$

following the same derivations as above. Therefore,

$$(z(S') - \varepsilon)\sqrt{2\delta\mu} \leq |\chi(S')| \leq (z(S') + \varepsilon)\sqrt{2\delta\mu}$$

for *any* convex set S' of A' (including A' itself), proving the first stipulation of 1.

For the second part of Theorem 1, consider a convex set $S \subset A^c$, with $z(S) > \varepsilon/16$ that lies completely inside a ball of radius r . For sufficiently large s , $\mathfrak{R}(S)$ will be completely disjoint from \mathfrak{P} ; however, $\mathfrak{D}(S)$ may not be. Thanks to the bounds on the measure of the outer and inner hulls,

$$|\mathfrak{D}(S)/\mathfrak{R}(S)| \leq \left(\frac{\varepsilon^2}{32} \right) m^d \tau \leq \left(\frac{\varepsilon^2}{26} \right) \tilde{\tau}_s.$$

To get an upper bound on $|\chi(S)|$, we assume that $\mathfrak{D}(S)$ has the maximal intersection with \mathfrak{P} , and that the remaining elements take on the maximal value allowed by Theorem 3. Specifically, this gives

$$\begin{aligned} |\chi(S)| &\leq |\chi(\mathfrak{U}(\mathfrak{D}(S)))| \\ &\leq |\mathfrak{D}(S) \cap \mathfrak{P}| \left(\frac{q}{\tilde{\tau}_s} \right) (1 + \tilde{\varepsilon}) + |\mathfrak{D}(S)| \left(\frac{\tilde{\varepsilon}q}{\tilde{\tau}_s} \right) \\ &\leq z(S) \left[\frac{\varepsilon^2}{26z(S)} (1 + \tilde{\varepsilon}) + K \left(1 + \frac{\varepsilon}{4} \right) 2\tilde{\varepsilon} \right] q \end{aligned}$$

using the earlier bound on the cardinality of the outer hull and the upper bound on the X_J , $J \notin \mathfrak{P}$ from Theorem 3. Thanks to the lower bound on

$z(S)$ and the definition of $\tilde{\varepsilon}$, the bracketed expression can be bounded above by $3\varepsilon/4$. Using the bound $q < (1 + \varepsilon/4)\sqrt{2\delta\mu}$ for all large s , derived above, and dividing through by $\sqrt{2\delta\mu}$, gives the desired result. This completes the proof of Theorem 1. \square

6. PROOF OF THEOREM 3

The bulk of the remaining length of the paper is dedicated to the proof of Theorem 3. The proof is organized as follows: First, we use concentration inequalities and large deviations estimate to show that, in the s -graded model conditioned to have too many edges, three inequalities hold with high probability. Next, we show that these inequalities imply the existence of a small set of indices such that almost all the “extra” edges have both endpoints in the union of its associated A_I ’s, and satisfies a convex upper bound. The next two subsections require delicate estimates which produce a subset which satisfies a similar quadratic lower bound, a termwise lower bound, and an additional upper bound on the total number of vertices in the set. Finally, simple convex analysis will show that this set will be a maximal clique set, and that the vertices are roughly equidistributed among the A_I ’s in this set.

6.1. Probabilistic Analysis. For this section, C will indicate any multiplicative constant that is uniformly bounded in n . It may depend on all other parameters, including s , and may take on different values from line to line.

In this section, we will estimate the probability of four rare events. First, we will produce a lower bound for the probability of the conditioning event – namely, that the number of edges exceeds its mean by a multiplicative factor of $(1 + \tilde{\delta})$. Next, we prove upper bounds for three events that will constrain the large deviation event. One will involve the number of I ’s for which X_I is big, the second will bound the distribution of vertices among the large X_I ’s, and the third will control the large deviation probability of the edge count between the small X_I ’s. We then combine all the estimates in the proof of 17, the essential probabilistic proposition of this paper.

Let L be the event $\{|E_s| > (1 + \tilde{\delta})\tilde{\mu}_s\}$. As before, let $q = (2\tilde{\delta}\tilde{\mu}_s)^{1/2}$ and $w = \tilde{\tau}_s\mathcal{D}$. Our first goal is to produce a lower bound on the probability of L :

Lemma 9. *If L , q and w are defined as above, then there exists a C such that*

$$\mathbb{P}(L) \geq \exp(-q(\log(q/w) - 1) - Cn^z \log n) ,$$

where $z = \max\{p/4, 3p/4 - 1/2\}$.

Proof. Pick a maximal clique set \mathfrak{K} and define the event H by

$$H := \left\{ \forall I \in \mathfrak{K}, X_I \geq \left\lceil \frac{q + n^z}{\tilde{\tau}_s} \right\rceil \right\} .$$

Note that the choice of z is somewhat arbitrary, and is motivated below; for now, it is important that, since $\log \tilde{\mu}_s \sim p \log n$ as $n \rightarrow \infty$ (by the definition of p), (3) implies that the correction term n^z is of lower order than q , because $\max\{p/4, 3p/4 - 1/2\} < p/2$ for all admissible values of p .

Assume that H occurs. This implies that the number of edges (in the graph G_s) with both endpoints in $\mathfrak{U}(\mathfrak{K})$ is at least

$$\binom{q + n^z}{2} \geq \delta \tilde{\mu}_s + \frac{qn^z}{2}$$

Therefore, if $H_{\mathfrak{K}}$ occurs then L would be true if the total number of edges with at most one endpoint in $\mathfrak{U}(\mathfrak{K})$ would exceed $\tilde{\mu}_s - (qn^z)/2$. Let $|E_s|'$ be the number of edges with neither endpoints in $\mathfrak{U}(\mathfrak{K})$. Then

$$\mathbb{E}(|E_s|') \geq \sum_{I: d(I, \mathfrak{K}) > s} \mathbb{E} \left[\binom{X_I}{2} + \frac{1}{2} \sum_{J: 0 < d(I, J) \leq s} X_I X_J \right].$$

This is a lower bound as we ignore pairs of nonzero elements which are close to \mathfrak{K} but not in \mathfrak{K} . The set $\{I : d(I, \mathfrak{K}) \leq s\}$ has diameter $\leq 3s$, and therefore its cardinality must be some constant independent of n . Therefore by (5),

$$\mathbb{E}(|E_s|') \geq (m^d - C) \frac{|N_I|n^2}{2m^{2d}} = \tilde{\mu}_s - \frac{C\tilde{\mu}_s}{m^d}.$$

We now wish to bound the probability

$$\mathbb{P}(|E_s|' < \tilde{\mu}_s - qn^z/2) \leq \mathbb{P} \left(|E_s|' < \mathbb{E}(|E_s|') - qn^z/2 + \frac{C\tilde{\mu}_s}{m^d} \right),$$

where the inequality follows from the above relation between the mean of $|E_s|'$ and $\tilde{\mu}_s$. By Lemma 4, $\tilde{\mu}_s/m^d$ grows as $n^{2(p-1)}$ (on the logarithmic scale). If $p < 1$, this is a vanishing quantity, and therefore

$$qn^z/2 - \frac{C\tilde{\mu}_s}{m^d} \leq qn^z/4.$$

If $p \geq 1$, then qn^z grows as $n^{5p/4-1/2}$. Since $p < 2 - \delta^*$ by the restrictions on r , $2(p-1) < 5p/4 - 1/2$, and, for sufficiently large n , the inequality above still holds.

Using the sum definition of $|E_s|$, a straightforward calculation can show that

$$\text{Var}(|E_s|') \leq \text{Var}(|E_s|) \leq Cm^d (\mathcal{D}^3 + \mathcal{D}^2).$$

Referring back to the growth bounds on m^d and \mathcal{D} , we see that the variance grows as n^p if $p < 1$ and n^{2p-1} if $p \geq 1$. By Chebyshev's inequality,

$$\mathbb{P}(|E_s|' < \mathbb{E}(|E_s|') - qn^z/4) \leq \frac{16\text{Var}(|E_s|')}{(qn^z)^2}.$$

For all value of p of interest, this fraction vanishes as $n^{-p/2}$, up to possible logarithmic factors. Thus, the probability of the event

$$\mathbb{P}(|E_s|' \geq \tilde{\mu}_s - qn^z/2) > 1 - \varepsilon$$

for any ε positive, for sufficiently large n .

Since the above event implies L once H occurs, this allows us to conclude that

$$\mathbb{P}(L) \geq \mathbb{P}(H)\mathbb{P}(L \mid H) \geq (1 - \varepsilon)\mathbb{P}(H).$$

Thus, the lower bound on the probability of L will come from a good lower bound on the probability of H . By definition,

$$\begin{aligned} \mathbb{P}(H) &= \left[\mathbb{P} \left(X_I \geq \left\lceil \frac{q + n^z}{\tilde{\tau}_s} \right\rceil \right) \right]^{\tilde{\tau}_s} \\ &\geq \left[\mathbb{P} \left(X_I = \left\lceil \frac{q + n^z}{\tilde{\tau}_s} \right\rceil \right) \right]^{\tilde{\tau}_s} \\ &\geq \left(e^{-\mathcal{D}} \mathcal{D}^{(q+2n^z)/\tilde{\tau}_s} / [(q + 2n^z)/\tilde{\tau}_s]! \right)^{\tilde{\tau}_s} \end{aligned}$$

where the final inequality follows by removing the ceiling function and compensating by adding a factor of two to the correction term. Now, by Stirling's approximation,

$$([(q + 2n^z)/\tilde{\tau}_s]!)^{-1} \geq \exp \left(- \left(\frac{q + 3n^z}{\tilde{\tau}_s} \right) \left[\log \left(\frac{q + 3n^z}{\tilde{\tau}_s} \right) - 1 \right] \right),$$

where the changed constant in front of the n^z term is to compensate for the omitted polynomial term in the approximation. Substituting this above gives

$$\mathbb{P}(H) \geq \exp \left(-\tilde{\tau}_s \mathcal{D} - (q + 3n^z) \left[\log \left(\frac{q + 3n^z}{\tilde{\tau}_s \mathcal{D}} \right) - 1 \right] \right).$$

Since n^z and $\tilde{\tau}_s \mathcal{D}$ are both lower order correction, we can find an absolute constant C such that the desired bound holds, completing the proof. \square

Let $a = \delta^*/25$ and $M = \max\{\mathcal{D}n^a, n^a\}$. We will say an index $I \in T$ is in the *bulk* if

$$X_I \leq M.$$

First, we wish to show that not too many indices are outside the bulk.

Lemma 10. *Let*

$$\alpha = \min\{1 - p/2 - a/2, p/2 - a/2\}$$

and define A to be the event

$$\{\exists S \subset T, |S| > n^\alpha \text{ such that } \forall I \in S, X_I > M\}.$$

Then, for some $C > 0$,

$$\mathbb{P}(A) \leq \exp(-Cn^{p/2+a/2})$$

Proof. Note that it is sufficient to show that no such set of cardinality exactly $\lceil n^\alpha \rceil$ exists. We can bound the probability of this event in a straightforward

manner. First, we establish the classical large deviation bound for Poisson random variables. To do so, note that

$$\mathbb{E}[\exp(\lambda X_I)] = e^{-\mathcal{D}} \sum_{k=0}^{\infty} \frac{\exp(\lambda k) \mathcal{D}^k}{k!} = \exp \left[\mathcal{D} (e^\lambda - 1) \right]$$

By Chebyshev's inequality, this implies that, for any positive λ ,

$$\mathbb{P}(X_I > t) \leq \exp \left[\mathcal{D} (e^\lambda - 1) - \lambda t \right]$$

Setting $\lambda = \log(t/\mathcal{D})$ gives

$$(7) \quad \mathbb{P}(X_I > t) \leq \exp(-t[\log(t/\mathcal{D}) - 1] - \mathcal{D}), \quad t > \mathcal{D}.$$

A similar procedure, using function $\mathbb{E}[e^{-\lambda X_I}]$, leads to

$$(8) \quad \mathbb{P}(X_I < t) \leq \exp(-t[\log(t/\mathcal{D}) - 1] - \mathcal{D}), \quad t < \mathcal{D}.$$

We now return to bounding the probability of A . If $p \geq 1$, we get that

$$(9) \quad \mathbb{P}(A) \leq \binom{m^d}{n^\alpha} \mathbb{P}(X_I > M)^{n^\alpha} \leq m^{dn^\alpha} \exp(-n^\alpha M \log(M/\mathcal{D})),$$

The polynomial rate of growth of $n^{\alpha+a}\mathcal{D}$ is $p/2 + a/2$, whereas n^α grows as $1-p/2-a/2$, which is smaller for all $p \geq 1$. Therefore, replacing $M \log(M/\mathcal{D})$ by $n^a\mathcal{D}$, which is strictly smaller, implies that

$$\mathbb{P}(A) \leq \exp(-n^{\alpha+a}\mathcal{D} + dn^\alpha \log m) \leq \exp(-Cn^{p/2+a/2})$$

If $p < 1$, the same calculation gives

$$(10) \quad \mathbb{P}(A) \leq \binom{m^d}{n^{p/2-a/2}} \exp(-n^{p/2+a/2}) \leq \exp(-Cn^{p/2+a/2}).$$

□

Next, we want some control over the behavior of indices outside the bulk. Define

$$Y_I = X_I (\log(X_I/\mathcal{D}) - 1) + \mathcal{D},$$

with the convention that $0 \cdot \log 0 = 0$. Note that $Y_I = \mathcal{I}(X_I)$, where \mathcal{I} is the rate function of a Poisson random variable of mean \mathcal{D} . Therefore, $\mathbb{P}[Y_I > t]$ should vanish as $\exp(-t)$, by “inverting” the rate function. We formalize this notion in the lemma below:

Lemma 11. *For any \mathcal{D} , any positive $\lambda < 1$, and some C uniform in n ,*

$$\mathbb{E}[\exp(\lambda Y_I)] \leq \frac{C}{1 - \exp(C(\lambda - 1)/\log \mathcal{D})}$$

Proof. First note that $Y_I \geq 0$, as the function $f(x) = x[\log(x/\mathcal{D}) - 1] + \mathcal{D}$ achieves its minimum at $x = \mathcal{D}$. This means that, in general, $f(x)$ is not invertible for all positive integer values of x . Therefore, we define two inverses. First, let

$$g_1(x) : [0, \mathcal{D}] \rightarrow [0, \mathcal{D}] \text{ be a function such that } (f \circ g_1)(x) = x.$$

Note that this function is decreasing, with $g_1(0) = \mathcal{D}$ and $g_1(\mathcal{D}) = 0$. For any $x > \mathcal{D}$, we say that $g_1(x) = -\infty$. We define g_2 , the second inverse, similarly, except its range is defined to be (\mathcal{D}, ∞) . This inverse is strictly increasing. We use (8) and (7) to deduce

$$\mathbb{P}(Y_I > t) = \mathbb{P}(X_I < g_1(t)) + \mathbb{P}(X_I > g_2(t)) \leq 2e^{-t}.$$

Note that, if $t > \mathcal{D}$, the first term contributes nothing to the probability.

These estimates allow us to bound the moment generating function of Y_I : let \mathcal{A} be the set of atoms of Y_I . Then, for any $\lambda < 1$,

$$\mathbb{E}[\exp(\lambda Y_I)] \leq \sum_{i \in \mathcal{A}} 2e^{-i+\lambda i}.$$

If $p < 1$, \mathcal{A} is more sparse than the integers, in the sense that $|\mathcal{A} \cap [0, n]| \leq n$. Thus, we can increase the expectation by taking the sum over the integers, and possibly adding a multiplicative constant. If $\mathcal{D} \rightarrow \infty$ (which would require $p \geq 1$), the minimal distance between adjacent atoms in \mathcal{A} is $C/\log \mathcal{D}$. Using these results, we conclude that, if $p < 1$,

$$\mathbb{E}[\exp(\lambda Y_I)] \leq C \sum_{i=1}^{\infty} e^{(\lambda-1)i} = \frac{C}{1 - \exp(\lambda - 1)}.$$

Meanwhile, when $p \geq 1$, can only produce the inferior bound

$$\mathbb{E}[\exp(\lambda Y_I)] \leq C \sum_{i=1}^{\infty} e^{C(\lambda-1)i/\log \mathcal{D}} \leq \frac{C}{1 - \exp(C(\lambda - 1)/\log \mathcal{D})}.$$

The second bound, which holds even if $p < 1$, is the one we required in the lemma. \square

The sum of the Y_I 's over indices *not* in the bulk will play a critical role in later analysis. We next produce an upper bound on the probability that there exists a small set of indices for which the sum of Y_I 's is particularly large. Intuitively, 10 shows that the complement of the bulk is not particularly large, so this estimate will control the desired quantity. Furthermore, the exponential tail of Y_I allows us to produce a very good upper bound on that sum.

Lemma 12. *Let Y_I and α as above, and define $\beta := p/2 - a/4$. Define the event*

$$B := \left\{ \exists S \subset T, |S| \leq n^\alpha \text{ such that } \sum_{I \in S} Y_I > t \right\},$$

where

$$t = q(\log(q/w) - 1) + n^\beta.$$

Then

$$\mathbb{P}(B) \leq \exp(-t + n^\beta/2).$$

Proof. We now use a Chernoff-like strategy to bound the probability of the event

$$\left\{ \sum_{I \in S} Y_I > t \right\}$$

for some $S \subset T$ of cardinality at most n^α . By Chebyshev's inequality

$$\begin{aligned} \mathbb{P}\left(\sum_{I \in S} Y_I > t\right) &\leq \frac{(\exp(\lambda Y_I))^{n^\alpha}}{\exp(\lambda t)} \\ &\leq \left(\frac{C}{1 - \exp(C(\lambda - 1)/\log \mathcal{D})} \right)^{n^\alpha} e^{-\lambda t}. \end{aligned}$$

Note that we used the estimate in 11 in that calculation.

We now set $\lambda = 1 - n^\alpha/t$. This leaves an exponential in $n^\alpha/(t \log \mathcal{D})$, which is vanishing in n . Therefore, we can bound

$$\begin{aligned} (1 - \exp(-Cn^\alpha/(t \log \mathcal{D})))^{-1} &\leq (Cn^\alpha/(t \log \mathcal{D}) - (Cn^\alpha/(t \log \mathcal{D}))^2)^{-1} \\ &\leq (Ct \log \mathcal{D})/n^\alpha + 2 \end{aligned}$$

for all sufficiently large n . Therefore,

$$\mathbb{P}\left(\sum_{I \in S} Y_I > t\right) \leq (C(t \log \mathcal{D})/n^\alpha + 2)^{n^\alpha} \cdot e^{-t+n^\alpha} \leq \exp(-t + Cn^\alpha \log n).$$

To calculate the probability of the event B defined in the lemma, we invoke the union bound:

$$\mathbb{P}(B) \leq \binom{m^d}{n^\alpha} \exp(-t + Cn^\alpha \log n).$$

We bound the combinatorial factor

$$\binom{m^d}{n^\alpha} \leq m^{dn^\alpha} \leq \exp^{Cn^\alpha \log n},$$

for some C independent of n , since m^d is a polynomial in n . Since $\beta > \alpha$, we can be sure that $Cn^\alpha \log n < n^\beta/2$ for all sufficiently large values of n . Therefore,

$$\mathbb{P}(B) \leq \exp(-t + n^\beta/2)$$

as required. \square

The final probabilistic step is to show that, conditional on L , the bulk does not contain very many extra edges. Let

$$\hat{X}_I := X_I \cdot 1_{X_I \leq M}$$

and $|\hat{E}_s|$ be the define analogously with $|E_s|$ by replacing X_I with its truncated version (Recall that $M = \max\{n^a, \mathcal{D}n^a\}$). In other words, $|\hat{E}_s|$ is the version of G_s obtained after deleting all vertices in boxes that satisfy $X_I > M$. Fix $\gamma = p - 2a$, and consider the event

$$D = \{|\hat{E}_s| - \tilde{\mu}_s > n^\gamma\}.$$

The argument for controlling the probability of D splits into two regimes. We first treat the situation in which \mathcal{D} is vanishing as n grows; this covers the case $p < 1$ and a subset of the possible values of $p = 1$. We wish to prove the following lemma:

Lemma 13. *Let D be as above, and assume that \mathcal{D} is vanishing as n grows. Then, for some $C > 0$,*

$$\mathbb{P}(D) \leq \exp(-Cn^{2\gamma-p-8a})$$

Proof. With our additional assumption of \mathcal{D} , we know that $M = n^a$. Therefore, $|\hat{E}_s|$ has a Lipschitz constant of Cn^{2a} in all its coordinates with respect to the Hamming metric when seen as a function of the \hat{X}_I 's.

We use Talagrand's convex concentration inequality [28, Theorem 4.1.1]. First, let us define the setting: let $\Omega = \prod_{i=1}^N \Omega_i$, where Ω_i are all probability spaces, the measure on Ω is the product measure, and X is a random variable. For a set $A \subset \Omega$, define the set

$$U_A(x) := \{\{s_i\} \in \{0, 1\}^N : \exists y \in A, s_i = 0 \implies x_i = y_i\}.$$

Let $V_A(x)$ be the convex hull of $U_A(x)$, and $d_c(A, x)$ is the ℓ^2 distance of $V_A(x)$ to the origin. For any set A , we denote A_t be the t blowup of A with respect to this metric, i.e.

$$A_t := \{x \in \Omega : d_c(A, x) \leq t\}.$$

We can now state the inequality:

Theorem 14 (Talagrand's Inequality). *If Ω , $\mathbb{P}[\cdot]$, A and A_t are as above, then*

$$\mathbb{P}[A] (1 - \mathbb{P}[A_t]) \leq e^{-t^2/4}.$$

We will not apply this theorem directly; instead, we use a corollary of this theorem frequently used in discrete settings [1, Theorem 7.7.1]. To do so, we consider f , a function from the natural numbers to the natural numbers. We say that f is a witness function for X if, whenever $X(\omega) \geq t$, there exists $I \subset [n]$ with $|I| \leq f(t)$, such that every ω' that agrees with ω in all $i \in I$ has $X(\omega') \geq t$. Furthermore, we assume that $X(\omega)$ is K -Lipschitz with respect to the Hamming distance — that is, $|X(\omega) - X(\omega')| \leq K$ whenever ω and ω' differ in at most one coordinate.

Theorem 15. *Let Ω be a product space, and X a real valued function on Ω with Lipschitz constant K with respect to the Hamming distance. If f is witness function for X as above, then, for any b and t ,*

$$\mathbb{P}[X > b + tK\sqrt{f(b)}] \mathbb{P}[X \leq b] \leq \exp(-t^2/4).$$

We now apply this theorem on $X = |\hat{E}_s|$. Since each coordinate is bounded above by n^a , X is Lipschitz with $K = Cn^{2a}$. The function $f(t) = 2t$ is a witness function for $|\hat{E}_s|$; to see this, note that $|\hat{E}_s|$ is the edge count of the s -graded geometric random graph, after we remove any X_I with very

high density. As such, we can “witness” the existence of t edges by finding at most $2t$ vertices; the flexibility of the setup allows us to pick these vertices judiciously, avoiding all the isolated ones. Finding $2t$ vertices will require at most $2t$ distinct coordinates, if each one of them vertices lies in a distinct A_I .

We apply the theorem with $b = \tilde{\mu}_s + qn^z$ and

$$t = \frac{n^\gamma - qn^z}{n^{2a}\sqrt{\tilde{\mu}_s + qn^z}},$$

with z defined as in the beginning of the section, and deduce that

$$\begin{aligned} \mathbb{P}(D)\mathbb{P}(|\hat{E}_s| \leq \tilde{\mu}_s + qn^z) &\leq \exp\left(-\frac{C[n^\gamma - qn^z]^2}{(\tilde{\mu}_s + qn^z)n^{4a}}\right) \\ &\leq \exp(-C'n^{2\gamma-p-8a}), \end{aligned}$$

where the final inequality follows because qn^z is a lower order correction to $\tilde{\mu}_s$, by definition of z .

Note that is inequality is only possible when $f(\tilde{\mu}_s) \leq m^d$; $f(\tilde{\mu}_s)$ represents the number of coordinates necessary to witness the property $|\hat{E}_s|$, so it must be bounded by the total number of random variables available. Since \mathcal{D} is vanishing, this is guaranteed to hold for all sufficiently large values of n .

Now, we must bound the probability that $|\hat{E}_s|$ is below $\tilde{\mu}_s + qn^z$. This is not a rare event: the mean of $|\hat{E}_s|$ is strictly smaller than the mean of $|E_s|$, and its variance is essentially the same. To see this second fact, note that

$$\text{Var}[|\hat{E}_s|] \leq \mathbb{E}\left[(|\hat{E}_s| - \tilde{\mu}_s)^2\right],$$

since the function $t \rightarrow \mathbb{E}[(X - t)^2]$ is minimized at the mean of X for any random variable. On the event $\{X_I = \hat{X}_I, \forall I \in T\}$, we can repeat the computation used in 9 to show that the variance is bounded above by a constant multiple of $\text{Var}[|E_s|]$. Meanwhile, thanks to the union bound, the complement occurs with probability of at most $m^d \exp(-Cn^a \log n)$. Since the maximal value of $(|\hat{E}_s| - \tilde{\mu}_s)^2$ is $m^{2d}n^{4a}$ by the coordinate-wise bound, we deduce that

$$\text{Var}[|\hat{E}_s|] \leq \text{Var}[|E_s|] + m^{3d}n^{4a} \exp(-Cn^a \log n) \leq Cm^d \mathcal{D}^2,$$

where the final inequality holds for all sufficiently large values of n , as in the computation in 9. Repeating the Chebyshev argument used in that proof, we can deduce that, for any $\varepsilon > 0$, there exists a sufficiently large n to ensure

$$\mathbb{P}(|\hat{E}_s| \leq \tilde{\mu}_s + qn^z) > 1 - \varepsilon.$$

This immediately implies

$$\mathbb{P}(D) \leq \exp(-Cn^{2\gamma-p-8a}),$$

completing the proof. \square

For the case \mathcal{D} being uniformly bounded from below, we use a different set of tools. Instead of Talagrand's inequality, the main tool of this proof is the Azuma-Hoeffding inequality. Unfortunately, if \mathcal{D} grows with n , a naive application of the inequality does not yield a sufficiently good bound for our purposes; the Lipschitz constant of $|\hat{E}_s|$, which is at most $C\mathcal{D}^2n^{2a}$, is too large. To get around this, we create a larger probability space, and calculate probabilities on this space. We will only apply the Azuma-Hoeffding inequality after we eliminate some pathological behavior by other means. Formally, we prove the lemma:

Lemma 16. *If \mathcal{D} is bounded above and below uniformly in n , then there exists a constant C' such that*

$$\mathbb{P}(D) \leq \exp\left(-\frac{C'n^{2\gamma}}{m^d n^{4a}}\right).$$

If \mathcal{D} grows as n grows, then

$$\mathbb{P}(D) \leq \exp\left(-\frac{C'n^{2\gamma}}{m^d \mathcal{D}^3 n^{4a}}\right).$$

Proof. We first consider the case in which \mathcal{D} is bounded. Then $|\hat{E}_s|$ has Lipschitz constant of at most $C\mathcal{D}^2n^{2a}$. By a naive application of Azuma-Hoeffding, we conclude that

$$\begin{aligned} \mathbb{P}(D) &\leq \mathbb{P}\left(|\hat{E}_s| - \mathbb{E}[|\hat{E}_s|] > n^\gamma\right) \\ &\leq \exp\left(-\frac{Cn^{2\gamma}}{2m^d \mathcal{D}^4 n^{4a}}\right) \\ &\leq \exp\left(-\frac{C'n^{2\gamma}}{m^d n^{4a}}\right), \end{aligned}$$

where the first inequality follows from the fact that $\tilde{\mu}_s > \mathbb{E}[|\hat{E}_s|]$, the second is Azuma-Hoeffding, and the final follows because \mathcal{D} is uniformly bounded in n . Note that, if \mathcal{D} was growing with n , the probability will vary with the *fourth* power of \mathcal{D} , and not the third power, as required by the lemma.

We now turn to the final case. The measure of each A_I is $1/m^d = \mathcal{D}/n$, which is much greater than $1/n$ by assumption on \mathcal{D} . We partition A_I into many sets of measure $1/n$; formally, let $\{F_{I,t}\}$, for natural $t \leq \lceil \mathcal{D} \rceil$ be a set of disjoint subsets of A_I such that $\lambda(F_{I,t}) = 1/n$ for every $t \leq \lceil \mathcal{D} \rceil - 1$, and

$$\bigcup_t F_{I,t} = A_I.$$

Note that the measure of the final $F_{I,t}$ will be strictly smaller than $1/n$, unless \mathcal{D} is an integer. We define $W_{I,t} = |\chi(F_{I,t})|$.

Clearly, $\sum_t W_{I,t} = X_I$. Define $\overline{|E_s|}$ as (yet another!) truncation of $|E_s|$ - this time, we truncate each $W_{I,t}$ at $n^a/2$. Formally, let $\overline{W_{I,t}} = W_{I,t} \cdot 1_{W_{I,t} < n^a/2}$, and define $\overline{|E_s|}$ by replacing each X_I in the definition

of $|E_s|$ by $\sum_t \overline{W_{I,t}}$. Every configuration of this truncation is also allowed in $|\hat{E}_s|$, but $|\overline{E_s}|$ has a far smaller Lipschitz constant with respect to the $W_{I,t}$'s. Specifically, each $W_{I,t}$ interacts with at most $|N_I|(\mathcal{D}+1)$ different variables, each one at most $n^a/2$. Thus, the Lipschitz constant is at most $C\mathcal{D} \cdot n^{2a}$ in the Hamming distance in the larger space. Of course, we pay two prices for this better truncation: first, the splitting procedure increases the dimensionality of the probability space. More importantly, we must control both $|\overline{E_s}|$ and $|\hat{E}_s| - |\overline{E_s}|$ to get control of the probability of D . The first task is a straightforward application of Azuma-Hoeffding:

$$\begin{aligned} \mathbb{P}\left[|\overline{E_s}| - \tilde{\mu}_s > n^\gamma/2\right] &\leq \mathbb{P}\left[|\overline{E_s}| - \mathbb{E}[|\overline{E_s}|] > n^\gamma/2\right] \\ &\leq \exp\left(-\frac{Cn^{2\gamma}}{m^d \mathcal{D}^3 n^{4a}}\right), \end{aligned}$$

using the fact that there are fewer than $m^d(\mathcal{D}+1)$ coordinates, each with Lipschitz constant $C\mathcal{D} \cdot n^{2a}$. This bound is sufficiently good for our purposes, as the denominator varies with third power of \mathcal{D} .

By partitioning,

$$\mathbb{P}[D] \leq \mathbb{P}\left[|\overline{E_s}| - \tilde{\mu}_s > n^\gamma/2\right] + \mathbb{P}\left[D, |\overline{E_s}| - \tilde{\mu}_s < n^\gamma/2\right].$$

The second event on the righthand side is dominated by the event $G := \{|\hat{E}_s| - |\overline{E_s}| > n^\gamma/2\}$. We now turn to bounding the probability of this event. The difference between the two random variables is given by configurations in which at least $W_{I,t}$ is larger than $n^a/2$. In fact,

$$|\hat{E}_s| - |\overline{E_s}| < \sum_{(I,t)} \left[\left(W_{I,t} \cdot 1_{W_{I,t} > n^a/2} \right) \cdot \sum_{J: d(I,J) \leq s} (X_J \cdot 1_{X_J < \mathcal{D}n^a}) \right].$$

While the random variables in the expression above are far from independent, we can replace the second sum over the X_J 's by $|N_I|\mathcal{D}n^a$, the upper bound imposed on it by the indicator random variables involved. Therefore,

$$\mathbb{P}[G] \leq \mathbb{P}\left[\sum_{(I,t)} \left(W_{I,t} \cdot 1_{W_{I,t} > n^a/2} \right) > \frac{Cn^\gamma}{\mathcal{D}n^a} \right]$$

To bound this final probability, we can directly bound the moment generating function of $W_{I,t} \cdot 1_{W_{I,t} > n^a/2}$:

$$\mathbb{E}\left[\exp\left(W_{I,t} \cdot 1_{W_{I,t} > n^a/2}\right)\right] \leq 1 + \sum_{k > n^a/2} \frac{e^k}{k!} \leq 1 + \exp(-n^a).$$

The first inequality follows because $W_{I,t}$ is a Poisson random variable of mean 1 (or possibly less than 1, if we pick the small $W_{I,t}$ in each I), while the second can be deduced by using Stirling's approximation and explicitly summing.

Applying a Chernoff strategy, we find that

$$\mathbb{P} \left[\sum_{(I,t)} \left(W_{I,t} \cdot 1_{W_{I,t} > n^a/2} \right) > \frac{Cn^\gamma}{\mathcal{D}n^a} \right] \leq (1 + \exp(-n^a))^{m^d(\mathcal{D}+1)} \cdot \exp \left(-\frac{Cn^\gamma}{\mathcal{D}n^a} \right).$$

Using the standard approximation $(1+x) \leq e^x$, we find that the prefactor is bounded by 2 for all n sufficiently large. Combining all the earlier estimates, we find that

$$\mathbb{P}[D] \leq \exp \left(-\frac{Cn^{2\gamma}}{m^d \mathcal{D}^3 n^{4a}} \right) + 2 \exp \left(-\frac{Cn^\gamma}{\mathcal{D}n^a} \right).$$

The first term vanishes like $\exp(-n^{1-8a}f(n))$, whereas the second vanishes as $\exp(-n^{1-3a}f'(n))$, with f and f' varying more slowly than any polynomial or rational function. Thus, the first term dominates, and gives us the desired upper bound on the probability of D . \square

We now combine the technical lemmas of this section into a single, essential proposition:

Proposition 17. *Let A , B , D , and L be defined as above. Then, in the conditional model, $\mathbb{P}(\cdot \mid L)$, the event $L \cap A^c \cap B^c \cap D^c$ occurs with high probability.*

Proof. It is sufficient to show that the sum $\mathbb{P}(L^c|L) + \mathbb{P}(A|L) + \mathbb{P}(B|L) + \mathbb{P}(D|L)$ vanishes as n grows. The first term is trivially equal to zero. To calculate the remaining three terms, we use the following obvious but very useful inequality:

$$\mathbb{P}(A|L) = \frac{\mathbb{P}(A \cap L)}{\mathbb{P}(L)} \leq \frac{\mathbb{P}(A)}{\mathbb{P}(L)}.$$

By 9 and 10, we know that

$$\mathbb{P}(A|L) \leq \exp \left(q[\log(q/w) - 1] + Cn^z \log n - C'n^{p/2+a/2} \right).$$

Since a is positive, the negative term dominates, and the probability vanishes as n grows. Referring to 12, we use the same technique to find that

$$\mathbb{P}(B|L) \leq \exp \left(Cn^z \log n - n^\beta/2 \right).$$

Since $\beta > z$ by construction, this probability vanishes as well. 13 guarantees that, when \mathcal{D} is vanishing,

$$\mathbb{P}(D|L) \leq \exp \left(q[\log(q/w) - 1] + Cn^z \log n - n^{2\gamma-p-8a} \right).$$

The negative term grows as n^{p-12a} , while the positive one grows as $n^{p/2}$ (both up to polynomial factors). Since $p > a/25$, the exponent of the negative term is always greater, and thus the probability vanishes. Using the definition of \mathcal{D} , m^d , and 16, we can show that, for any \mathcal{D} bounded below,

$$\mathbb{P}(D|L) \leq \exp \left(q[\log(q/w) - 1] + Cn^z \log n - n^{1-8a} \cdot f(n) \right),$$

where $f(n)$ grows or vanishes slower than any polynomial. As long as $p < 2 - 16a$ (which is guaranteed by our assumptions on p), this probability vanishes, and the proof of the proposition is complete. \square

6.2. Showing Partial Localization. Surprisingly, the proposition that ends the previous subsection actually implies Theorem 3, in the sense that *any* configuration in $L \cap A^c \cap B^c \cap D^c$ contains a maximal clique set such that the X_I associated with every index is arbitrarily close to $(q/\tilde{\tau}_s)$. To prove this rather counterintuitive fact, we now embark on a technical proof which relies on careful asymptotics and convex analysis – and no probabilistic analysis.

For the remains of the proof of 3, all statements hold conditional on $L \cap A^c \cap B^c \cap D^c$, and, therefore, with high probability in the conditional model.

We now define some functionals of subsets of T , which will ease the notation. For any $W \subset T$, let

$$Q(W) := \frac{2}{q^2} \sum_{I \in W} \left(\binom{X_I}{2} + \frac{1}{2} \sum_{J \in N_I \cap W} X_I X_J \right).$$

This is the number of edges with both endpoints in $\mathfrak{U}(W)$, normalized by $q^2/2$. With this normalization, showing that most extra edges have both endpoints in $\mathfrak{U}(W)$ is equivalent to showing that $Q(W)$ is bounded below by something close to 1. Similarly, for two disjoint index sets W and W' , we let

$$Q(W, W') := \frac{2}{q^2} \sum_{I \in W} \sum_{J \in N_I \cap W'} X_I X_J.$$

This counts the edges between W and W' with the same normalization as above. Note that, for any $W \subset T$,

$$|E_s| = \frac{q^2}{2} (Q(W) + Q(W, W^c) + Q(W^c)).$$

Next, define

$$V(W) := \frac{1}{q} \sum_{I \in W} X_I,$$

the number of vertices in W . With this normalization, we can see that $V(W)^2 \geq Q(W)$. We also normalize the cardinality of sets by $\tilde{\tau}_s$, that is, for any $W \subset T$,

$$h(W) := \frac{|W|}{\tilde{\tau}_s}.$$

As a final preliminary, we phrase Jensen's inequality, an essential tool of this analysis, in terms of the quantities we defined:

Lemma 18. *For any $W \subset T$,*

$$\frac{1}{q} \sum_{I \in W} Y_I \geq V(W) \left[\log \left(\frac{q}{w} \right) + \log V(W) - \log h(W) - 1 \right]$$

Proof. This is a direct application of Jensen's inequality to the convex function $Y_I = X_I(\log(X_I/\mathcal{D}) - 1) + \mathcal{D}$. \square

We now consider the set of indices whose associated X_I 's are much larger than expected. The next lemma proves three bounds on this set:

Lemma 19. *Define \mathfrak{I} by*

$$\{I \in T : X_I > \max\{n^a, \mathcal{D} \cdot n^a\}.$$

Then, conditional on $L \cap A^c \cap B^c \cap D^c$, for any $\xi > 0$ and n sufficiently large (depending only on ξ),

$$(11) \quad \frac{1}{q} \sum_{I \in \mathfrak{I}} Y_I \leq \log(q/w) - 1 + \xi,$$

$$(12) \quad V(\mathfrak{I}) \leq C,$$

$$(13) \quad Q(\mathfrak{I}) \geq 1 - \frac{\xi}{\log n}.$$

Note that the lower bound on $Q(\mathfrak{I})$ ensures that there are $\tilde{\delta}\tilde{\mu}_s$ edges - i.e. all the “excess” forced by the conditioning on L - with both endpoints in $\mathfrak{U}(\mathfrak{I})$, up to lower order corrections. While the upper bound on $V(\mathfrak{I})$ is very loose, the upper bound on $\sum Y_I$ is both tight and essential to our analysis.

Proof. Defining \mathfrak{I} as in the lemma above, we know that, thanks to A^c ,

$$(14) \quad |\mathfrak{I}| \leq n^\alpha.$$

With this cardinality bound, we know that \mathfrak{I} falls within the definition of B^c , and therefore

$$\sum_{I \in \mathfrak{I}} Y_I \leq q(\log(q/w) - 1) + n^\beta/2.$$

Fix $\xi > 0$. Dividing through by q and noting that n^β/q vanishes as n grows, we can increase n to ensure that (11) holds, satisfying the first stipulation.

We now apply 18 to this inequality. We deduce that

$$V(\mathfrak{I}) \left[\log \left(\frac{q}{wh(\mathfrak{I})} \right) - 1 \right] + V(\mathfrak{I}) \log V(\mathfrak{I}) \leq \log(q/w) - 1 + \xi.$$

If $V(\mathfrak{I}) \leq 1$, we have a (much better than needed!) bound on $V(\mathfrak{I})$. Otherwise, $V(\mathfrak{I}) \log V(\mathfrak{I})$ is positive, and we conclude that

$$V(\mathfrak{I}) \leq \frac{\log(q/w) - 1 + Cn^\beta/q}{\log(q/(wh(\mathfrak{I}))) - 1}.$$

By (14), and the definitions of the variables q , w and α ,

$$\frac{q}{wh(\mathfrak{I})} \geq Cn^{a/2},$$

for some constant C ; therefore the denominator grows at least as a constant multiple of $\log n$. Meanwhile, $q/w \leq Cn^{1-p/2}$ for some constant C , meaning that

$$V(\mathfrak{I}) \leq \frac{2(1-p/2)\log n}{(a/4)\log n} \leq \frac{8(1-p/2)}{a}.$$

We set the righthand side equal to C ; since it is independent of n , the second stipulation has been proved.

For the final stipulation, we use the event $L \cap D^c$. By D^c ,

$$(15) \quad Q(\mathfrak{I}^c) - \mu \leq n^\gamma.$$

By the occurrence of L ,

$$\frac{q^2}{2}(Q(\mathfrak{I}) + Q(\mathfrak{I}, \mathfrak{I}^c) + Q(\mathfrak{I}^c)) \geq (1 + \tilde{\delta})\tilde{\mu}_s$$

Assuming $Q(\mathfrak{I}^c)$ takes on the maximal value allowed by (15), we deduce that

$$(16) \quad Q(\mathfrak{I}) + Q(\mathfrak{I}, \mathfrak{I}^c) \geq 1 - \frac{2n^\gamma}{q^2}.$$

Next, we bound $Q(\mathfrak{I}, \mathfrak{I}^c)$ by replacing each term in the sum over the neighborhood N_I by the maximum allowable value:

$$(17) \quad Q(\mathfrak{I}, \mathfrak{I}^c) \leq \frac{2}{q^2} \sum_{I \in \mathfrak{I}} X_I \left(\sum_{J \in N_I} \max_{J \in \mathfrak{I}^c} X_J \right) \leq \frac{2|N_I|M}{q} \left(\frac{1}{q} \sum_{I \in \mathfrak{I}} X_I \right).$$

The parenthetical expression is $V(\mathfrak{I})$, which is at most a constant independent of n . By 4, $|N_I|$ is also bounded above by a constant, and thus

$$Q(\mathfrak{I}, \mathfrak{I}^c) \leq \frac{CM}{q} \leq \frac{\xi}{2 \log n},$$

for sufficiently large values of n , since M/q vanishes polynomially fast in n for any value of p allowed. Similarly, we can increase n sufficiently to ensure that $(2n^\gamma)/q$ is also bounded above by $\xi/(2 \log n)$ (which is possible thanks to the definition of γ). Substituting the bounds into (16) completes the proof. \square

6.3. Improving the Linear Bound. Our goal is to produce an upper bound on the linear sum that matches the convex lower bound already in place for \mathfrak{I} . To do so, we find a subset of \mathfrak{I} that has a similar number of “internal” edges, but which only involves approximately q vertices. Using the convex upper bound from above, we construct such a subset in the simplest way possible: taking the largest elements of \mathfrak{I} , and stopping when the total number of vertices gets close to q . However, although we hope that the cardinality of this set has a uniform upper bound in n , our argument is too rough to prove such a statement. Therefore, we ensure that the bounds get tighter as the size of the set increases; later these bounds will allow us to enforce an upper bound on the number of indices we are concerned with.

Lemma 20. *Let \mathfrak{J} be the set defined in equation 19. Then there exists $\mathfrak{T} \subset \mathfrak{J}$ such that*

$$Q(\mathfrak{T}) \geq 1 - \psi(\mathfrak{T})$$

and

$$1 - \frac{2\xi}{\log n} \leq V(\mathfrak{T}) \leq 1 + \phi(\mathfrak{T})$$

where

$$\phi(\mathfrak{T}) = \min \left\{ \xi/3, \frac{2}{|\mathfrak{T}|} - \frac{2\xi}{\log n} \right\}$$

and

$$\psi(\mathfrak{T}) = \min \left\{ \xi, \frac{C'}{|\mathfrak{T}|} \right\}$$

for some C' independent of n .

Proof. Since $Q(\mathfrak{J}) \geq 1 - \xi/\log n$ by (13), we know that

$$V(\mathfrak{J}) \geq \sqrt{1 - \xi/\log n} \geq 1 - \xi/\log n$$

for all sufficiently large n . Now, we order the elements of \mathfrak{J} by size; for the purposes of this proof, we label the X_I 's by integers: let X_1 the largest X_I among all $I \in \mathfrak{J}$, X_2 be the second largest, and so on. Pick an integer k , and define \mathfrak{J}_k to be the first k terms of \mathfrak{J} with respect to this order. Finally, define

$$(18) \quad \mathfrak{T} := \text{The first } \mathfrak{J}_k \text{ such that } V(\mathfrak{J}_k) > 1 - 2\xi/\log n.$$

that is, the smallest subset of \mathfrak{J} that has at least $q(1 - 2\xi/\log n)$ vertices of χ in its associated A_I 's. By minimality,

$$V(\mathfrak{T} \setminus \{X_{|\mathfrak{T}|}\}) \leq 1 - \frac{2\xi}{\log n}$$

Furthermore, $X_{|\mathfrak{T}|}$ must be the minimal element of \mathfrak{T} , and therefore,

$$\frac{X_{|\mathfrak{T}|}}{q} \leq \frac{V(\mathfrak{T})}{|\mathfrak{T}|}.$$

Substituting these two inequalities into the identity $V(\mathfrak{T}) = V(\mathfrak{T} \setminus \{X_{|\mathfrak{T}|}\}) + X_{|\mathfrak{T}|}/q$ and solving for $V(\mathfrak{T})$ gives

$$V(\mathfrak{T}) \leq 1 + \frac{2}{|\mathfrak{T}|} - \frac{2\xi}{\log n}.$$

Note that this inequality may be derived from the previous one only if $|\mathfrak{T}|$ is larger than a universal constant. That, however, is not a problem; Corollary 21 at the end of the proof shows that $|\mathfrak{T}|$ must be large enough, simply using the definition of \mathfrak{T} and Lemma 18.

This establishes one half of the desired upper bound on $V(\mathfrak{T})$. It remains to show that $V(\mathfrak{T})$ is bounded by ξ when $|\mathfrak{T}| < 6/\xi$. Assume that that bound, and apply Lemma 18 to \mathfrak{T} :

$$\begin{aligned} & V(\mathfrak{T}) \left[\log \left(\frac{q}{w} \right) - \log h(\mathfrak{T}) - 1 \right] + V(\mathfrak{T}) \log V(\mathfrak{T}) \\ & \leq \frac{1}{q} \sum_{I \in \mathfrak{T}} Y_I \leq \log(q/w) - 1 + \xi, \end{aligned}$$

where the last upper bound follows by (11) and the fact that $\mathfrak{T} \subset \mathfrak{J}$. From the above inequality it follows that either $V(\mathfrak{T}) \leq 1$, or

$$V(\mathfrak{T}) \leq \frac{\log(q/w) - 1 + \xi}{\log(q/w) - \log h(\mathfrak{T}) - 1}.$$

The assumed upper bound on $|\mathfrak{T}|$ implies that $h(\mathfrak{T})$ is uniformly bounded; since $|\log(q/w)| \rightarrow \infty$ as $n \rightarrow \infty$, we conclude the second upper bound on $V(\mathfrak{T})$.

Next, we need to show that $Q(\mathfrak{T})$ is large. In the remaining part of this proof, \mathfrak{T}^c will denote the complement of \mathfrak{T} in \mathfrak{J} , and not in all of T . It is sufficient to bound

$$Q(\mathfrak{T}, \mathfrak{T}^c) + Q(\mathfrak{T}^c)$$

from above, by partitioning the sum that defines $Q(\mathfrak{J})$, as in the previous section. The maximum element in \mathfrak{T}^c cannot exceed $(qV(\mathfrak{T}))/|\mathfrak{T}|$, as it must be smaller than the minimum of the elements in \mathfrak{T} , by definition. Therefore, for any $I \in \mathfrak{J}$,

$$\sum_{J \in N_I \cap \mathfrak{T}^c} X_J \leq |N_I| \max_{J \in \mathfrak{T}^c} X_J \leq |N_I| \cdot \left(\frac{qV(\mathfrak{T})}{|\mathfrak{T}|} \right).$$

Expanding the definition of both terms, we see that

$$(19) \quad Q(\mathfrak{T}, \mathfrak{T}^c) + Q(\mathfrak{T}^c) \leq \frac{4}{q^2} \sum_{I \in \mathfrak{J}} X_I \sum_{J \in N_I \cap \mathfrak{T}^c} X_J \leq [V(\mathfrak{J})V(\mathfrak{T})] \frac{|N_I|}{|\mathfrak{T}|}.$$

Setting $C' = 2C|N_I|$, where C is the constant from 19 and assuming $|\mathfrak{T}| > C'/\xi$, we see that this

$$Q(\mathfrak{T}, \mathfrak{T}^c) + Q(\mathfrak{T}^c) \leq \psi(\mathfrak{T}),$$

noting that $V(\mathfrak{T}) \leq 1 + \phi(\mathfrak{T}) \leq 2$.

Finally, assume that $|\mathfrak{T}| \leq C'/\xi$. We go back to the upper bound on Y_I 's in \mathfrak{J} :

$$\frac{1}{q} \left(\sum_{I \in \mathfrak{T}} Y_I + \sum_{J \in \mathfrak{T}^c} Y_J \right) \leq \log(q/w) - 1 + \xi.$$

We apply Lemma 18 to the term involving \mathfrak{T} in the above inequality to conclude that

$$\begin{aligned} & V(\mathfrak{T}) \left[\log \left(\frac{q}{w} \right) + \log V(\mathfrak{T}) - \log h(\mathfrak{T}) - 1 \right] + \frac{1}{q} \sum_{J \in \mathfrak{T}^c} Y_J \\ & \leq \log(q/w) - 1 + \xi. \end{aligned}$$

Next, we substitute the lower bound on $V(\mathfrak{T})$ (from definition). After some algebraic manipulation, we conclude that

$$\begin{aligned} \frac{1}{q} \sum_{J \in \mathfrak{T}^c} Y_J & \leq \xi + \frac{3\xi}{\log n} + \log h(\mathfrak{T}) \\ & \quad + \frac{2\xi}{\log n} \left[\log \left(\frac{q}{w} \right) - \frac{3\xi}{\log n} - \log h(\mathfrak{T}) - 1 \right]. \end{aligned}$$

(To get this, we bounded $\log(1 - 2\xi/\log n)$ from below by $-3\xi/\log n$, an estimate that holds for all sufficiently large n .) Thanks to the assumed upper bound on $|\mathfrak{T}|$, the right-hand side can be bounded above by $2\log(1/\xi)$ for all sufficiently small ξ and sufficiently large n , as all terms except $h(\mathfrak{T})$ are bounded above by a constant multiple of ξ .

With this upper bound, we can now apply Lemma 18 to \mathfrak{T}^c to conclude that

$$V(\mathfrak{T}^c) \left[\log \left(\frac{q}{wh(\mathfrak{T}^c)} \right) - 1 \right] + V(\mathfrak{T}^c) \log V(\mathfrak{T}^c) \leq 2\log(1/\xi).$$

Since the absolute minimum of the function $x \log x$ is $-1/e$, we can replace $V(\mathfrak{T}^c) \log V(\mathfrak{T}^c)$ by this minimum and maintain the lower bound. We also know that $wh(\mathfrak{T}^c)$ is bounded above by $\mathcal{D}n^\alpha$ (since $\mathfrak{T}^c \subset \mathfrak{I}$ and the event A^c is assumed to have happened). Recalling the definition of α , we now arrive at the inequality

$$V(\mathfrak{T}^c) \leq \frac{2\log(1/\xi) + 1/e}{\log(q/(\mathcal{D}n^\alpha)) - 1} \leq \frac{3\log(1/\xi)}{a/2\log n} \leq \frac{\xi}{C},$$

where C is, again, the constant from 19. The final inequality follows since the numerator is bounded in n , while the denominator grows, and therefore the fraction can be made smaller than any constant. From here, a bound on $Q(\mathfrak{T}, \mathfrak{T}^c) + Q(\mathfrak{T}^c)$ is easy: For any $I \in \mathfrak{I}$,

$$\frac{1}{q} \sum_{J \in N_I \cap \mathfrak{T}^c} X_J \leq V(\mathfrak{T}^c).$$

Therefore,

$$Q(\mathfrak{T}, \mathfrak{T}^c) + Q(\mathfrak{T}^c) \leq \frac{4}{q^2} \sum_{I \in \mathfrak{I}} X_I \sum_{J \in N_I \cap \mathfrak{T}^c} X_J \leq \left(\frac{\xi}{C} \right) V(\mathfrak{I}) \leq \xi$$

Completing the proof of the lemma. \square

Corollary 21. *There exists C independent of n such that*

$$|\mathfrak{T}| > \tilde{\tau}_s.$$

Proof. Since $\mathfrak{T} \subset \mathfrak{J}$, (11) applies to \mathfrak{T} as well. We now apply Lemma 18 to this set, and conclude that

$$V(\mathfrak{T}) \left[\log \left(\frac{q}{w} \right) + \log V(\mathfrak{T}) - \log h(\mathfrak{T}) - 1 \right] \leq \log \left(\frac{q}{w} \right) + \xi - 1$$

From the definition of \mathfrak{T} in the proof of Lemma 20, we have a lower bound on $V(\mathfrak{T})$. Noting that

$$\log \left(1 - \frac{2\xi}{\log n} \right) \geq \frac{-4\xi}{\log n}$$

for all sufficiently large n , we can conclude that

$$- \left(1 - \frac{2\xi}{\log n} \right) \log(h(\mathfrak{T})) \leq \xi + \frac{2\xi \log(q/w)}{\log n} + \frac{C\xi}{\log n}$$

We recall that $q/w \leq Cn^{1-p/2}$, and therefore there exists a constant C such that

$$-\log(h(\mathfrak{T})) \leq C\xi.$$

Inverting the negative logarithm gives

$$h(\mathfrak{T}) \geq \exp(-C\xi) \geq 1 - C\xi$$

If we decrease ξ enough to ensure that $C\xi < \tilde{\tau}_s/2$ (which is possible, as $\tilde{\tau}_s$ is independent of n), we find that

$$|\mathfrak{T}| \geq \tau_s - 1/2.$$

Since the cardinality is an integer, we get the desired bound. \square

The estimates of 20 imply that $V(\mathfrak{T})$ satisfies nearly matching upper and lower bounds; we now exploit this property to bound an algebraic quantity that has a nice geometric interpretation:

Proposition 22. *For any $W \subset T$, define*

$$P_I(W) := \frac{1}{q} \sum_{J \in W, d(I,J) > s} X_J.$$

Note that the sum is over indices whose distance exceeds s . Then

$$\frac{1}{q} \sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{T}) \leq 3\phi(\mathfrak{T}) + \psi(\mathfrak{T}),$$

where \mathfrak{T} , $\phi(\cdot)$ and $\psi(\cdot)$ are defined as in 20.

Proof. First, we consider the quantity $[V(\mathfrak{T})]^2 - Q(\mathfrak{T})$. By expanding the product and cancelling, we can deduce that

$$[V(\mathfrak{T})]^2 - Q(\mathfrak{T}) = \frac{2}{q} \sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{T}) + \frac{V(\mathfrak{T})}{q}.$$

Note that

$$X_I^2 - 2 \binom{X_I}{2} = X_I,$$

which accounts for the additive factor of $V(\mathfrak{T})/q$. By 20, we conclude that

$$\frac{1}{q} \sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{T}) \leq ([V(\mathfrak{T})]^2 - Q(\mathfrak{T})) \leq (1 + \phi(\mathfrak{T}))^2 - (1 - \psi(\mathfrak{T})) \leq 3\phi(\mathfrak{T}) + \psi(\mathfrak{T}).$$

□

The quantity $\sum X_I P_I(\mathfrak{T})$ is commensurate with the number of edges in the complement graph of $G_s(\chi, r)$, restricted to vertices in $\mathfrak{U}(\mathfrak{T})$. We wish to show that, whenever it is small, the graph restricted to \mathfrak{T} is not far from a clique. This will be the goal of the next section.

6.4. Removing Lower Order Terms. Before proceeding, consider the case $\xi = 0$. In this case, both ψ and ϕ vanish. Since all terms in lefthand sum are strictly positive (in fact, $\mathfrak{T} \subset \mathfrak{I}$, so every X_I we consider is at least M), $P_I(\mathfrak{T})$ would be forced to vanish for each $I \in \mathfrak{T}$. This would imply that the diameter of \mathfrak{T} is at most s . By 21, the cardinality of the set is at least $\tilde{\tau}_s$ – and thus \mathfrak{T} would be a maximal clique set!

Since ξ is strictly positive, we cannot apply this argument to the set \mathfrak{T} . The next lemma produces a subset of \mathfrak{T} with slightly stronger bounds; at the end of the section, we will apply an argument very similar to the one described above to prove that that set is a maximal clique set.

Lemma 23. *Let \mathfrak{T} be as above, and define \mathfrak{P} such that*

$$\mathfrak{P} := \left\{ I \in \mathfrak{T} : X_I > \frac{\xi^{1/4} q}{\tilde{\tau}_s} \right\}$$

Then, for sufficiently small $\xi > 0$ and sufficiently large n ,

$$1 - \xi^{1/5} \leq V(\mathfrak{P}) \leq 1 + \phi(\mathfrak{T})$$

and

$$\frac{1}{q} \sum_{I \in \mathfrak{P}} Y_I \leq V(\mathfrak{P})(\log(q/w) - 1) + 2\xi^{1/10}.$$

Proof. We split \mathfrak{T} into three sets: \mathfrak{P} is defined as above, while the rest of the indices are split as follows:

$$\mathfrak{L}_1 := \left\{ I \in \mathfrak{T} : X_I \leq \frac{\xi^{1/4} q}{\tilde{\tau}_s \log |\mathfrak{T}|} \right\}$$

and

$$\mathfrak{L}_2 := \left\{ I \in \mathfrak{T} : \frac{\xi^{1/4} q}{\tilde{\tau}_s \log |\mathfrak{T}|} \leq X_I \leq \frac{\xi^{1/4} q}{\tilde{\tau}_s} \right\}.$$

Since the three sets partition \mathfrak{T} , we can prove the lemma as with sufficiently good upper bounds on $V(\mathfrak{L}_i)$ and lower bounds on the sum of the Y_I 's in both sets.

To bound $V(\mathfrak{L}_1)$, we first need to bound $P_I(\mathfrak{L}_1)$ from below. The worst case scenario is that the distance restriction removes the $|N_I|$ largest elements of \mathfrak{L}_1 . Therefore,

$$P_I(\mathfrak{L}_1) \geq V(\mathfrak{L}_1) - \frac{1}{q}|N_I| \left(\max_{J \in \mathfrak{L}_1} X_J \right) \geq V(\mathfrak{L}_1) - \frac{|N_I|\xi^{1/4}}{\tilde{\tau}_s \log |\mathfrak{T}|}.$$

Since $P_I(W) \leq P_I(W')$ whenever $W \subset W'$, we see that 22 implies that

$$\frac{1}{q} \sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{L}_1) \leq 3\phi(\mathfrak{T}) + \psi(\mathfrak{T}).$$

Replacing $P_I(\mathfrak{L}_1)$ with its minimum and recalling that $|N_I|/\tilde{\tau}_s$ is uniformly upper bounded (by Lemma 4), we see that

$$\left(V(\mathfrak{L}_1) - \frac{C\xi^{1/4}}{\log |\mathfrak{T}|} \right) V(\mathfrak{T}) \leq 3\phi(\mathfrak{T}) + \psi(\mathfrak{T}).$$

Using the (very suboptimal) lower bound of $1/2$ for $V(\mathfrak{T})$, we conclude that

$$(20) \quad V(\mathfrak{L}_1) \leq 6\phi(\mathfrak{T}) + 2\psi(\mathfrak{T}) + \frac{C\xi^{1/4}}{\log |\mathfrak{T}|}.$$

for some C independent of n . Repeating this analysis with \mathfrak{L}_2 yields the inequality

$$V(\mathfrak{L}_2) \leq 6\phi(\mathfrak{T}) + 2\psi(\mathfrak{T}) + C\xi^{1/4}.$$

Since both ϕ and ψ are bounded above by ξ , we get

$$(21) \quad \max\{V(\mathfrak{L}_1), V(\mathfrak{L}_2)\} \leq \xi^{1/5}/3.$$

Now,

$$V(\mathfrak{P}) = V(\mathfrak{T}) - V(\mathfrak{L}_1) - V(\mathfrak{L}_2) \geq 1 - \frac{2\xi}{\log n} - \frac{2\xi^{1/5}}{3} \geq 1 - \xi^{1/5}$$

This establishes the lower bound on $V(\mathfrak{P})$. The upper bound follows trivially from $\mathfrak{P} \subset \mathfrak{T}$.

Finally, we need to improve the upper bound on the Y_I 's associated with \mathfrak{P} . Not surprisingly, we will use Jensen's inequality to get lower bounds on the Y_I associated with \mathfrak{L}_1 and \mathfrak{L}_2 . By inclusion, we know that

$$\frac{1}{q} \sum_{I \in \mathfrak{P}} Y_I \leq (\log(q/w) - 1 + \xi) - \sum_{I \in \mathfrak{L}_1 \cup \mathfrak{L}_2} Y_I$$

Suppose that

$$(22) \quad \frac{1}{q} \sum_{I \in \mathfrak{L}_i} Y_I \geq V(\mathfrak{L}_i) (\log(q/w) - 1) - \xi^{1/10}/2$$

for $i = 1, 2$. The inequality

$$1 - V(\mathfrak{L}_1) - V(\mathfrak{L}_2) \leq V(\mathfrak{P}) + (2\xi)/\log n$$

follows from partitioning \mathfrak{T} into its three constituent sets, and the lower bound on $V(\mathfrak{T})$ from Lemma 20. Substituting the two inequalities into the earlier statement gives

$$\begin{aligned} \frac{1}{q} \sum_{I \in \mathfrak{P}} Y_I &\leq V(\mathfrak{P}) (\log(q/w) - 1) + \xi + \frac{2\xi}{\log n} (\log(q/w) - 1) + \xi^{1/10} \\ &\leq V(\mathfrak{P}) (\log(q/w) - 1) + 2\xi^{1/10}. \end{aligned}$$

using the fact that $\log(q/w)$ is bounded above by a constant multiple $\log n$. Since this is the required inequality in the statement of the lemma, we will be done if we can prove the inequality (22).

By applying Jensen's inequality 18 to \mathfrak{L}_i , we know that

$$(23) \quad \frac{1}{q} \sum_{I \in \mathfrak{L}_i} Y_I \geq V(\mathfrak{L}_i) (\log(q/w) - 1) + V(\mathfrak{L}_i) \log V(\mathfrak{L}_i) - V(\mathfrak{L}_i) \log h(\mathfrak{L}_i).$$

We have already shown in (21) that $V(\mathfrak{L}_i)$ is small. Now, the function $\xi^a \log \xi$ is bounded below $-\xi^b/4$ for any $b < a$ and ξ sufficiently close to zero, and therefore,

$$V(\mathfrak{L}_i) \log V(\mathfrak{L}_i) \geq -\xi^{1/10}/4.$$

For the final term, we must employ a different strategy for \mathfrak{L}_1 and \mathfrak{L}_2 . If $|\mathfrak{L}_1| \leq C/\xi$, then $-V(\mathfrak{L}_1) \log h(\mathfrak{L}_1) > C\xi^{1/5} \log \xi$. Repeating the argument from before, the final term can be bounded below by $-\xi^{1/10}/4$, as required. Now suppose $|\mathfrak{L}_1| > C/\xi$. In this range, both $\phi(\mathfrak{T})$ and $\psi(\mathfrak{T})$ are bounded above by $C/|\mathfrak{T}|$. Thus, (20) implies that

$$V(\mathfrak{L}_1) \leq \frac{\xi^{1/5}}{\log |\mathfrak{T}|}.$$

Since $\mathfrak{L}_1 \subset \mathfrak{T}$, we deduce that $h(\mathfrak{L}_1) \leq C|\mathfrak{T}|$, and

$$-V(\mathfrak{L}_i) \log h(\mathfrak{L}_i) > -C\xi^{1/5} > -\xi^{1/10}/4.$$

This completes the lemma for \mathfrak{L}_1 .

For \mathfrak{L}_2 , we claim that $|\mathfrak{L}_2| < C/\xi^{1/4}$. To see this, first assume that $P_I(\mathfrak{L}_2) = 0$ for some I . This implies that $|\mathfrak{L}_2| \leq |N_I|$, and $h(\mathfrak{L}_2)$ is bounded above by a constant by 4. Now, assume that $P_I(\mathfrak{L}_2)$ is nonzero for all I . Then we can minimize the value of P_I by assuming that the sum only includes the minimal elements of \mathfrak{L}_2 :

$$P_I(\mathfrak{L}_2) \geq \left(\frac{|\mathfrak{L}_2| - |N_I|}{q} \right) \cdot \left(\min_{J \in \mathfrak{L}_2} X_J \right) \geq \frac{C\xi^{1/4}|\mathfrak{L}_2| - |N_I|}{\log |\mathfrak{T}|},$$

for some C independent of n , where the second inequality follows from the definition of \mathfrak{L}_2 (and using 4 to push the factors of $\tilde{\tau}_s$ into C). Multiplying

by X_I/q , summing over I 's in \mathfrak{T} , and using 22, we see that

$$\begin{aligned} V(\mathfrak{T}) \cdot \left(\frac{C\xi^{1/4}|\mathfrak{L}_2| - |N_I|}{\log |\mathfrak{T}|} \right) &\leq \frac{1}{q} \sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{L}_1) \\ &\leq 3\phi(\mathfrak{T}) + \psi(\mathfrak{T}). \end{aligned}$$

Solving for $|\mathfrak{L}_2|$, we deduce that

$$|\mathfrak{L}_2| \leq |N_I| + \frac{C \log |\mathfrak{T}| (\phi(\mathfrak{T}) + \psi(\mathfrak{T}))}{\xi^{1/4}},$$

where we use the upper bound on $V(\mathfrak{T})$. Since both $\phi(\mathfrak{T})$ and $\psi(\mathfrak{T})$ are bounded above by $2/|\mathfrak{T}|$, the numerator is at most a constant independent of n . Therefore, $|\mathfrak{L}_2| < C/\xi^{1/4}$. To complete the proof, we simply repeat the argument used in the case $|\mathfrak{L}_1| \leq C/\xi$. \square

As promised, we now show that \mathfrak{P} has the desired geometric properties:

Corollary 24. *The set \mathfrak{P} is a maximal clique set.*

Proof. Assume that there exists a pair of indices $I, J \in \mathfrak{P}$ such that $d(I, J) > s$. By the definition of \mathfrak{P} ,

$$\frac{X_I X_J}{q^2} \geq \frac{\xi^{1/2}}{\tilde{\tau}_s^2}.$$

By 22, we know this cannot exceed $3\phi(\mathfrak{T}) + \psi(\mathfrak{T})$, which is bounded above by 4ξ . This forces

$$\frac{\xi^{1/2}}{\tilde{\tau}_s^2} \leq 4\xi.$$

However, the above inequality can easily be rendered false by choosing ξ so small that $\xi \leq (2\tilde{\tau}_s)^{-4}$. This can be done, because by Lemma 4, $\tilde{\tau}_s$ is uniformly bounded by a constant. With such a choice of ξ , the diameter of \mathfrak{P} is bounded above by s .

Combining Lemma 18 and 23 gives

$$\begin{aligned} V(\mathfrak{P}) (\log(q/w) - 1) + V(\mathfrak{P}) (\log V(\mathfrak{P}) - \log h(\mathfrak{P})) \\ \leq V(\mathfrak{P}) (\log(q/w) - 1) + 2\xi^{1/10}, \end{aligned}$$

and therefore

$$h(\mathfrak{P}) \geq V(\mathfrak{P}) \exp\left(-\frac{2\xi^{1/10}}{V(\mathfrak{P})}\right) \geq V(\mathfrak{P}) \left(1 - \frac{2\xi^{1/10}}{V(\mathfrak{P})}\right)$$

using the standard estimate $e^{-x} \geq 1 - x$. Combining this with the lower bound on $V(\mathfrak{P})$ from Lemma 23 forces

$$|\mathfrak{P}| = \tilde{\tau}_s h(\mathfrak{P}) \geq \tilde{\tau}_s (1 - 3\xi^{1/10})$$

Again, since ξ is under our control, we set $\xi < (2\tilde{\tau}_s)^{-10}$. This implies that

$$|\mathfrak{P}| \geq \tilde{\tau}_s - \frac{3}{2^{10}}$$

Cardinality must be an integer, and therefore $|\mathfrak{P}| \geq \tilde{\tau}_s$. Therefore, we have shown that \mathfrak{P} is a set of diameter at most s , with at least $\tilde{\tau}_s$ elements. But, by definition, $\tilde{\tau}_s$ is the largest possible cardinality of any set of diameter s . Thus, \mathfrak{P} must have diameter *exactly* s — otherwise, consider the union of \mathfrak{P} with one of the indices in T at distance 1 from \mathfrak{P} . By the triangle inequality, the diameter of this set is at most s , but its cardinality is $\tilde{\tau}_s + 1$, contradicting maximality. Thus, \mathfrak{P} is a maximal clique set. \square

6.5. Convex Analysis. We are nearly done with the proof: all that remains is to show that the elements of \mathfrak{P} are close to $q/\tilde{\tau}_s$, and to improve the upper bound for maximal clique sets far away from \mathfrak{P} .

Lemma 25. *Let \mathfrak{P} be as above. Then, for any I in \mathfrak{P} ,*

$$\left| \frac{X_I \tilde{\tau}_s}{q} - 1 \right| < \xi^{1/40}.$$

Proof. We consider the Taylor's expansion of Y_I around the value $q/\tilde{\tau}_s$. Explicitly, we let $f(x) = x(\log(x/\mathcal{D}) - 1) + \mathcal{D}$, and by Taylor's theorem,

$$Y_I = f\left(\frac{q}{\tilde{\tau}_s}\right) + f'\left(\frac{q}{\tilde{\tau}_s}\right)\left(X_I - \frac{q}{\tilde{\tau}_s}\right) + \frac{f''(L(X_I))}{2}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2$$

where $L(X_I)$ is some number between X_I and $q/\tilde{\tau}_s$. Differentiating $f(x)$ explicitly and simplifying algebraically, we see that

$$Y_I = \mathcal{D} - \frac{q}{\tilde{\tau}_s} + X_I \log(q/w) + \frac{1}{2L(X_I)}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2.$$

Next, we sum over \mathfrak{P} and use the upper bound from Lemma 23 to deduce that

$$\begin{aligned} & \frac{1}{q} \sum_{I \in \mathfrak{P}} \left[\mathcal{D} - \frac{q}{\tilde{\tau}_s} + X_I \log(q/w) + \frac{1}{2L(X_I)}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2 \right] \\ & \leq V(\mathfrak{P})(\log(q/w) - 1) + 2\xi^{1/10}. \end{aligned}$$

Recalling that \mathfrak{P} has exactly $\tilde{\tau}_s$ elements, a bit of algebraic manipulation yields

$$(24) \quad \frac{1}{q} \sum_{I \in \mathfrak{P}} \frac{1}{2L(X_I)}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2 \leq 1 - V(\mathfrak{P}) + 2\xi^{1/10} \leq 3\xi^{1/10},$$

where the final inequality follows thanks to the lower bound on $V(\mathfrak{P})$ from 23.

Now, define

$$\mathfrak{W}_1 := \left\{ I \in \mathfrak{P} : X_I \geq (1 + \xi^{1/40})q/\tilde{\tau}_s \right\}$$

and

$$\mathfrak{W}_2 := \left\{ I \in \mathfrak{P} : X_I \leq (1 - \xi^{1/40})q/\tilde{\tau}_s \right\}.$$

On \mathfrak{W}_1 , the function $1/L(X_I)$ is bounded below by $1/X_I$. Thus, (24) implies that

$$\frac{|\mathfrak{W}_1|}{q} \min_{I \in \mathfrak{W}_1} \left(\frac{1}{2X_I} \left(X_I - \frac{q}{\tilde{\tau}_s} \right)^2 \right) \leq 3\xi^{1/10}.$$

The function $x \mapsto (x - q/\tilde{\tau}_s)^2/(2x)$ is strictly increasing on the interval $[(q/\tilde{\tau}_s)(1 + \xi^{1/40}), \infty)$, and thus,

$$|\mathfrak{W}_1| \cdot \left(\frac{\xi^{1/20}}{2\tilde{\tau}_s \cdot (1 + \xi^{1/40})} \right) \leq 4\xi^{1/10}.$$

Reducing ξ such that $\xi < (10\tilde{\tau}_s)^{-20}$ implies that $|\mathfrak{W}_1|$ is bounded above by $4/5$; since the cardinality is an integer, the set is empty. For \mathfrak{W}_2 , we can bound $1/L(X_I)$ from below by $\tilde{\tau}_s/q$. Thus,

$$\frac{|\mathfrak{W}_2|}{q} \min_{I \in \mathfrak{W}_2} \left(\frac{\tilde{\tau}_s}{2q} \left(X_I - \frac{q}{\tilde{\tau}_s} \right)^2 \right) \geq \frac{|\mathfrak{W}_2|\xi^{1/20}}{2\tilde{\tau}_s}.$$

Substituting this into (24) and reducing ξ sufficiently completes the proof. \square

We have completed the proof of the difficult assertion in 3; all that is left is to ensure the second stipulation holds.

Proof of 3. Set $\xi = \min\{\tilde{\epsilon}^{40}, 1\}$. By the sequence of assertions 17, 19, 20, 23, 24 and 25, we produce a maximal clique set \mathfrak{P} such that

$$\left| \frac{X_I \tilde{\tau}_s}{q} - 1 \right| < \tilde{\epsilon},$$

proving the first stipulation of 3.

For the second stipulation, we must show that all elements of \mathfrak{P}^c are small. If $\mathfrak{T} \setminus \mathfrak{P}$ is nonempty, the largest element in that set must satisfy $X_I < \xi^{1/4} q / \tau_s$. Furthermore, every element in $\mathfrak{I} \setminus \mathfrak{T}$ is smaller than every element in \mathfrak{T} , implying this bound is inherited by all elements of $\mathfrak{I} \setminus \mathfrak{P}$. This is enough to pass the upper bound to all remaining X_I 's, as required.

We are left with the scenario in which $\mathfrak{P} = \mathfrak{T}$. Formally, it is still possible that the largest X_I with $I \in \mathfrak{I} \setminus \mathfrak{T}$ is of order q . However, we now know that $h(\mathfrak{T}) = 1$. By Lemma 18 and the lower bound in Lemma 20, we deduce that

$$\begin{aligned} \frac{1}{q} \sum_{I \in \mathfrak{T}} Y_I &\geq V(\mathfrak{T}) (\log(q/w) + \log(V(\mathfrak{T})) - 1) \\ &\geq \left(1 - \frac{2\xi}{\log n} \right) (\log(q/w) - 1). \end{aligned}$$

Therefore, we find that

$$\begin{aligned} \frac{1}{q} \sum_{I \in \mathfrak{I} \setminus \mathfrak{T}} Y_I &\leq [\log(q/w) - 1 + \xi] - \left(1 - \frac{2\xi}{\log n}\right) [\log(q/w) - 1] \\ &\leq \frac{4\xi \log(q/w)}{\log n} \\ &\leq C\xi. \end{aligned}$$

As before, the final inequality follows because $\log(q/w) \leq C \log n$ for some C .

If there is some $I^* \in \mathfrak{I} \setminus \mathfrak{T}$ with $X_{I^*} > \xi q$, we find that

$$\begin{aligned} Y_{I^*} &= X_{I^*} (\log(X_{I^*}/\mathcal{D}) - 1) + \mathcal{D} \\ &\geq \xi q (\log(q/\mathcal{D}) - 1), \end{aligned}$$

where we ignore several positive terms in the final inequality. If we divide through by q , we find that this expression still grows with n , whereas the earlier upper bound is uniformly bounded. This is a contradiction, and therefore every $X_I \in \mathfrak{I} \setminus \mathfrak{T}$ is bounded by ξq . This completes the proof. \square

7. PROOF OF THE UPPER TAIL LARGE DEVIATION PRINCIPLE

As usual, we write μ , τ , and r instead of μ_n , τ_n and r_n . We now prove Theorem 2, which claims that the function

$$I(x) := \left(\frac{2-p}{2}\right) \sqrt{2x}$$

is the upper tail rate function for the random variable $(|E| - \mu)/\mu$ with speed $\sqrt{\mu} \log n$. Recall that we restrict our attention to subsets of the interval $(0, \infty)$, as our result only holds for events in which $|E|$ exceeds its expectation.

Instead of proving 2 directly, we will prove the following proposition instead:

Proposition 26. *Let*

$$R_t := \left\{ \frac{|E| - \mu}{\mu} \geq t \right\}.$$

Then

$$(25) \quad \lim_{n \rightarrow \infty} \frac{\log \mathbb{P}[R_t]}{\sqrt{\mu} \log n} = -I(t).$$

This statement is standard, but its proof is straightforward and we include it for completeness.

Proof that 26 implies 2. Pick F to be a closed subset of $(0, \infty)$, and let a_F be its leftmost endpoint. Since $I(x)$ is increasing, its infimum over F occurs at a_F . Furthermore, $F \subset [a_F, \infty)$, and therefore,

$$\mathbb{P}\left[\frac{|E| - \mu}{\mu} \in F\right] \leq \mathbb{P}\left[\frac{|E|}{\mu} \in [a_F, \infty)\right] = \mathbb{P}[R_{a_F}].$$

Taking the logarithm, dividing by $\sqrt{\mu} \log n$, and applying (25) gives the upper bound for F .

Next, take G open and pick $b \in G$. For some $\varepsilon > 0$, $[b, b + \varepsilon) \in G$. Therefore,

$$\mathbb{P}\left[\frac{|E| - \mu}{\mu} \in G\right] \geq \mathbb{P}\left[\frac{|E| - \mu}{\mu} \in [b, b + \varepsilon)\right] = \mathbb{P}[R_b] - \mathbb{P}[R_{b+\varepsilon}].$$

Applying (25) twice, we deduce that, for any $\delta > 0$, there is an n sufficiently large to ensure that

$$\begin{aligned} \mathbb{P}\left[\frac{|E| - \mu}{\mu} \in G\right] &\geq \exp(-(1 + \delta) \cdot I(b) \cdot \sqrt{\mu} \log n) \\ &\quad - \exp(-(1 - \delta) \cdot I(b + \varepsilon) \cdot \sqrt{\mu} \log n). \end{aligned}$$

Picking δ sufficiently small (as a function of ε) ensures that the second term is smaller than half the first term. Taking logarithms and dividing by $\sqrt{\mu} \log n$ establishes the lower bound on the probability of $\{(|E| - \mu)/\mu \in G\}$, and establishes 2. \square

Proof of 26. We prove the estimate on the probability of R_t in two parts. First, we show the upper bound. Fix $\varepsilon > 0$. For an arbitrary pair of events A and B , assume that, conditional on A , the event B occurs with probability at least $1 - \varepsilon$. This implies that

$$\mathbb{P}[A] \leq \left(\frac{1}{1 - \varepsilon}\right) \mathbb{P}[B].$$

By Theorem 1, there exists a sufficient large n such that conditioning on R_t implies that the random geometric graph has a clique of size at least $\sqrt{2t\mu}(1 - \varepsilon)$ with probability at least $1 - \varepsilon$. This means that, for any s , the s graded model includes a maximal clique set $\mathfrak{P} \subset T$ with at least as many vertices as the clique of the random geometric graph. Since every maximal clique set has $\tilde{\tau}_s$ indices, there can be at most $m^{d\tilde{\tau}_s}$ distinct maximal clique sets; this is an egregious overcount, but we have no need for finer control. Thus, by the union bound, the probability that there exists a maximal clique set with $\sqrt{2t\mu}$ vertices is bounded above by $m^{d\tilde{\tau}_s}$ times the probability that a single one has the same property. The number of vertices in a maximal clique set is distributed as a Poisson random variable of mean $\tilde{\tau}_s \mathcal{D} = w$. Therefore, the chain of implication allows us to conclude that

$$\mathbb{P}[R_t] \leq \left(\frac{m^{d\tilde{\tau}_s}}{1 - \varepsilon}\right) \mathbb{P}\left[\text{Poisson}(w) > \sqrt{2t\mu}(1 - \varepsilon)\right].$$

Let $v := \sqrt{2t\mu}(1 - \varepsilon)$. Applying (7) to the right-hand side above gives

$$\begin{aligned} \mathbb{P}[R_t] &\leq \left(\frac{m^{d\tilde{\tau}_s}}{1 - \varepsilon} \right) \exp \left(-v \left[\log \left(\frac{v}{w} \right) - 1 \right] + w \right) \\ &\leq \exp \left(-(1 - 2\varepsilon) \sqrt{2t\mu} \log \left[\frac{\sqrt{\mu}}{w} \right] \right), \end{aligned}$$

where the second inequality follows for all sufficiently large n by noting that all the missing terms vanish in comparison to $\sqrt{\mu} \log n$, and can therefore be absorbed at the cost of changing ε to 2ε . By the definitions of μ , p and w ,

$$\frac{\sqrt{\mu}}{w} = n^{(2-p)/2} h(n),$$

for some $h(n)$ varying more slowly than any polynomial. Therefore,

$$\frac{1}{\sqrt{\mu} \log n} \log \mathbb{P}[R_{t,n}] \leq -(1 - 2\varepsilon) \left(\frac{2-p}{2} \right) \sqrt{2t} + \frac{\log h(n)}{\log n}.$$

Since ε is arbitrary, we conclude the desired upper bound.

For the lower bound, we replicate the argument found in the proof of 9; unfortunately, the s -graded model approximates the edge count of the random geometric graph from above, not from below, and so we cannot simply apply that lemma directly to deduce the lower bound. Specifically, we find a configuration which implies R_t . Fix B , a ball of diameter r , and let H' be the event that there are $\sqrt{2t\mu} + n^z$ vertices in B , where z is defined as in the beginning of Section 6.1. Since the number of vertices in B is a Poisson random variable of mean $n\tau$, we can explicitly compute that

$$\begin{aligned} \mathbb{P}[H'] &= \mathbb{P}[\text{Poisson}(n\tau) > \sqrt{2t\mu} + n^z] \\ &\geq \mathbb{P}[\text{Poisson}(n\tau) = \lceil \sqrt{2t\mu} + n^z \rceil]. \end{aligned}$$

As discussed in Section 6.1, this shows that

$$\mathbb{P}[H'] \geq \exp \left(-[\sqrt{2t\mu} + 3n^z] \log \left[\frac{\sqrt{2t\mu} + 3n^z}{n\tau} \right] \right).$$

Absorbing all terms that grow in n more slowly than $\sqrt{\mu} \log n$, we can deduce that, for sufficiently large n ,

$$(26) \quad \mathbb{P}[H'] \geq \exp(-(1 + \varepsilon)I(t)\sqrt{\mu} \log n),$$

where we bound the rate of growth of $\log[\sqrt{\mu}/(n\tau)]$ using the definitions of p and τ .

The event R_t will follow if the number of edges with at most one endpoint in B exceeds $\mu - 2\sqrt{2t\mu} \cdot n^z$. Let $|E'|$ be the number of edges with no endpoints in B . Letting $1_{i,j}$ be the indicator of an edge between vertices i

and j , we can see that

$$\begin{aligned}\mathbb{E}(|E'|) &= \mathbb{E} \left[\binom{N}{2} \mathbb{E}(1_{1,2} \cdot 1_{v_1, v_2 \notin B} \mid N) \right] \\ &= \frac{n^2}{2} \mathbb{P}(\{\|v_1 - v_2\| \leq r\} \cap \{v_1, v_2 \notin B\}),\end{aligned}$$

where N is the total number of point in the torus, as before, and the probability measure in the second equality is given by the uniform process. For notational convenience, let $1_{i,j}^B$ be the indicator of the event $\{\|v_i - v_j\| \leq r\} \cap \{v_i, v_j \notin B\}$, and μ^B be its expectation under the measure of the uniform process (by symmetry, this is independent of the indices i and j). If v_1 is at a distance greater than r from B , the second condition holds trivially. For a fixed B , the probability that v_1 is within distance r of B is a constant multiple of r^d . Thus,

$$\mu^B \geq (1 - Cr^d)\nu r^d,$$

for some C that depends only on the norm and the dimension. Thus, the expected value of $|E'|$ is bounded below by $\mu(1 - Cr^d)$. Therefore, by definition of z , we have the inequality

$$(27) \quad \mu - 2\sqrt{2t\mu} \cdot n^z \leq \mathbb{E}[|E'|] - \frac{\sqrt{2t\mu} \cdot n^z}{4}$$

We now need a variance estimate for $|E'|$:

$$\text{Var}(|E'|) = \mathbb{E}(\text{Var}(|E'| \mid N)) + \text{Var}(\mathbb{E}(|E'| \mid N)).$$

We have already calculated the expectation of $|E'|$ given N above; since μ^B does not depend on N , we deduce that

$$\text{Var}(\mathbb{E}(|E'| \mid N)) = (\mu^B)^2 \text{Var} \left[\binom{N}{2} \right].$$

A standard calculation will show that the variance of $\binom{N}{2}$ is $n^3 + n^2/2$. Meanwhile,

$$\mu^B \leq \mathbb{P}(\|v_1 - v_2\| \leq r) = \nu r^d.$$

Combining these facts gives

$$\text{Var}(\mathbb{E}(|E'| \mid N)) \leq Cr^{2d}n^3,$$

for some universal constant C .

Next, we estimate the expression $\mathbb{E}(\text{Var}(|E'| \mid N))$. We can write this variance as

$$\text{Var}(|E'| \mid N) = \mathbb{E} \left[\left(\sum_{1 \leq i < j \leq N} 1_{i,j}^B - \mu^B \right)^2 \mid N \right].$$

We now decompose this sum into three sums by distributing the square: one sum over pairs of the form $(i, j), (k, l)$ with four distinct indices, one with pairs of the form $(i, j), (i, k)$ where one index repeats, and the final over perfect squares of terms involving (i, j) . The expectation of the first one is

zero, as the event that i, j form an edge with both endpoints outside of B is completely independent of the same event occurring over distinct vertices k, l . For a fixed choice of (i, j) and (i, k) , we can bound

$$\mathbb{E}[1_{i,j}^B \cdot 1_{i,k}^B] \leq \mathbb{P}[\|v_i - v_j\| \leq r, \|v_i - v_k\| \leq r] = (\nu r^d)^2,$$

where the first inequality follows by removing the requirement that the vertices lie outside of B , and thus increasing the probability. There are $N(N-1)(N-2)$ ways to choose a pair of indices that overlap in exactly one entry. Thus,

$$\sum (1_{i,j}^B - \mu^B)(1_{i,k}^B - \mu^B) = \sum (1_{i,j}^B \cdot 1_{i,k}^B - (\mu^B)^2) \leq C' r^{2d} N^3,$$

for some universal constant C' . Again, this overestimates the value of this sum dramatically, but is sufficient for our purposes. Finally, the contribution of terms of the form $(1_{i,j}^B - \mu^B)^2$ to the sum is exactly $\binom{N}{2}(\mu_B - \mu_B^2)$, which is bounded above by $C'' r^d N^2$. Combining these results, taking expectations over N , and then adding the contribution of the variance of the expectation from before, we conclude that

$$(28) \quad \text{Var}(|E'|) \leq C'''(r^d n^2 + r^{2d} n^3),$$

for yet another universal constant C''' . $r^d n^2$ grows as n^p , while $r^{2d} n^3$ grows as n^{2p-1} - both up to factors that vanish in comparison to any polynomial. Thus, the variance of $|E'|$ is $n^p f(n)$ if $p \leq 1$, and $n^{2p-1} g(n)$ when $p > 1$, with $f(n)$ and $g(n)$ grow more slowly than any polynomial or rational function in n .

By Chebyshev's inequality and (27),

$$\begin{aligned} \mathbb{P}[|E'| < \mu - 2\sqrt{2t\mu} \cdot n^z] &\leq \mathbb{P}\left[|E'| < \mathbb{E}[|E'|] - \frac{\sqrt{2t\mu} \cdot n^z}{4}\right] \\ &\leq \frac{8\text{Var}(|E'|)}{t\mu \cdot n^{2z}}. \end{aligned}$$

Regardless of the value of p , this quantity grows as $n^{-p/2}$ up to logarithmic factors (just as it did in Section 6.1), and therefore, with probability $1 - \varepsilon$, $|E'|$ exceeds $\mu - 2\sqrt{2t\mu} \cdot n^z$ for all sufficiently large n . Following the chain of implication, we see that

$$\mathbb{P}[R_t] \geq \mathbb{P}[H'](1 - \varepsilon).$$

Substituting the earlier bound (26) on the probability of H' completes the proof. \square

Acknowledgment. The authors thank Mathew Penrose for many helpful comments. They would also like to thank the anonymous referees for their thorough reports, which greatly improved this manuscript.

REFERENCES

- [1] Alon, N. and Spencer, J. H. (2008) *The Probabilistic Method*. John Wiley and Sons, Hoboken, New Jersey.
- [2] Avin, C., and Ercal, G. (2007). On the cover time and mixing time of random geometric graphs. *Theoretical Computer Science*, 380(1), 2–22 .
- [3] Burago, Y. D. and Zalgaller, V. A. (1998). *Geometric Inequalities*. Leningrad.
- [4] Chatterjee, S. (2010). A note about the uniform distribution on the intersection of a simplex and a sphere. *Preprint*. Available at <http://arxiv.org/abs/1011.4043>.
- [5] Chatterjee, S. (2012). The missing log in large deviations for triangle counts. *Random Structures and Algorithms*, 40(4), 437–451.
- [6] Chatterjee, S. and Dembo, A. (2014). Nonlinear large deviations. *Preprint*. Available at <http://arxiv.org/abs/1401.3495>.
- [7] Demarco, B. and Kahn, J. (2012). Upper tails for triangles. *Random Structures and Algorithms*, 40(4), 452–459.
- [8] Dembo, A. and Zeitouni, O. (2010). *Large Deviations Techniques and Applications*. (Vol 38). Springer, New York.
- [9] Federer, H. (1996). *Geometric Measure Theory*. Springer, Berlin.
- [10] Gilbert, E. N. (1961). Random Plane Networks. *Journal of the Society of Industrial and Applied Mathematics*, 9(4), 533–543.
- [11] Goel, A., Rai, S., and Krishnamachari B. (2005). Monotone properties of random geometric graphs have sharp thresholds. *Annals of Applied Probability*, 15(4), 2535–2552.
- [12] Grimmett, G. (1999). *Percolation*. Springer, Berlin.
- [13] Hafner, R. (1972). The asymptotic distribution of clumping. *Computing*, 10(4), 335–351.
- [14] Janson, S. (2004). Large deviations for sums of partly dependent random variables. *Random Structures and Algorithms*, 24(3), 234–248.
- [15] Janson, S. and Rucinski, A. (2002). The infamous upper tail. *Random Structures and Algorithms*, 20(3), 317–342.
- [16] Janson, S., Oleszkiewicz, K., and Rucinski, A. (2004). Upper tails for subgraph counts in random graphs. *Israel Journal of Mathematics*, 142, 61–92.
- [17] Kim, J.H. and Vu, V.H. (2000). Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3), 417– 434.
- [18] Lubetzky, E. and Zhao, Y. (2014). On the variational problem for upper tails in sparse random graphs. *Preprint*. Available at <http://arxiv.org/abs/1402.6011>.
- [19] Meester, R. and Roy, R. (1996). *Continuum Percolation*. (Vol. 119) Cambridge University Press, Cambridge.
- [20] Muller, T. (2006). Two-point concentration in random geometric graphs. *Combinatorica*, 28(5), 529–545.
- [21] Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *The Annals of Probability* 7(5), 745–789.
- [22] Penrose, M. D. (2002). Focusing of the scan statistic and geometric clique number. *Advances in Applied Probability*, 739–753.
- [23] Penrose, M. (2003). *Random Geometric Graphs*. (Vol. 5) Oxford University Press, Oxford.
- [24] Penrose, M. D. and Yukich, J.E. (2003). Weak laws of large numbers in geometric probability. *The Annals of Applied Probability*, 13(1), 277–303.
- [25] Penrose, M. D. and Yukich, J. E. (2005). Normal approximation in geometric probability. *Steins Method and Applications, Lecture Note Series, Institute for Mathematical Sciences, National University of Singapore*, 5, 37–58.
- [26] Petersen, P. (2006). *Riemannian Geometry*. (Vol. 171). Springer, New York.

- [27] Schreiber, T. and Yukich, J. E. (2005). Large deviations for functionals of spatial point processes with applications to random packing and spatial graphs. *Stochastic Processes and Their Applications*, 115(8), 1332–1356.
- [28] Talagrand, M. (1996). Concentration of measures and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'I.H.E.S.*, 81(1), 73–205.
- [29] Talagrand, M. (2003). *Spin Glasses: A Challenge for Mathematicians*. Springer, Berlin.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
EMAIL: `souravc@stanford.edu`

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GENEVA
EMAIL: `Matan.Harel@unige.ch`